IDENTIFICATION OF PERTURBATION-RESPONSIVE KEY TRANSCRIPTION FACTORS IN TRANSCRIPTIONAL REGULATORY NETWORKS

by

Fatma Beste Kınıkoğlu

B.S. in Ch.E., Middle East Technical University, 2003

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2006

ACKNOWLEDGEMENTS

It is a pleasure to thank many people who made this thesis possible.

Most of all, I would like to deeply thank my thesis supervisor Prof. Betül Kırdar for her enthusiastic supervision. She provided me with lots of good ideas, invaluable insights and encouragement during my graduate study. Without her ideas, care and consideration, this thesis would not have matured.

I would also like to express my sincere gratitude to my thesis co-supervisor Prof. Zeynep İlsen Önsan for her help and support during the preparation of this dissertation.

I am grateful to the members of my thesis committee for devoting their precious time to read and comment on my thesis.

I would like to take this opportunity to thank all my fiends in our laboratory for their companionship throughout the overall process, Burcu Öztük, Duygu Dikicioğlu, Erdem Aygök, F. Betül Kavun, F. Özde Ütkür, Pınar Pir and Saliha Durmuş. I would like to thank particularly Tunahan Çakır and K. Yalçın Arga for kindly answering my questions and discussing my research with me.

Most importantly, I must express my special thanks to my friend Yener Yeşilırmak, whose contribution to my life, including my graduate study, is clearly priceless.

Finally, I am forever indebted to my parents, my grandmother and my brother Batu for providing me with hope, peace, support, inspiration, advices and most importantly with joy throughout my whole life. To them I dedicate this thesis.

This thesis has been supported by the project DPT-03K120250.

ABSTRACT

IDENTIFICATION OF PERTURBATION-RESPONSIVE KEY TRANSCRIPTION FACTORS IN TRANSCRIPTIONAL REGULATORY NETWORKS

In eukaryotic cells, most genes are found to be regulated by various temporary and permanent transcription factors whose activity levels change as response to perturbations. Discovering the underlying mechanism of complex cellular processes and responses to perturbations is a major challenge in post-genomic research. In this M.S. study, so-called 'key' transcription factors (transcription factors around which the most significant transcriptional changes occur) responding significantly to genetic and environmental perturbations were identified in yeast *Saccharomyces cerevisiae* by an algorithm based on hypothesis-driven data analysis. In contrast to existing approaches, the proposed approach uses network topology for determining the activity levels of transcription factors. The identification of the perturbation-responsive key transcription factors provides a dynamic perspective of transcriptional regulation which has central role in cellular function and structure. An extensive genome-scale map of transcriptional regulatory network in *S. cerevisiae* was constructed and integrated with gene expression data. The analysis of yeast data suggests that the method is capable of successfully identifying perturbation-responsive key transcription factors and it provides valuable information about transcription factors and their conditional/temporal behavior. In this study, it was also showed that once the key transcription factors are identified, the perturbation-responsive subnetworks might be revealed by interconnecting the key transcription factors and their target genes differentially expressed when the same perturbation is introduced. Furthermore, for each key transcription factor, its best candidate target genes were predicted (their regulatory interactions are not experimentally justified yet), which are differentially expressed after the same perturbations and their promoter regions contain bindig site(s) for the key transcription factor.

ÖZET

Ökaryot hücrelerde, çoğu genin ekspresyonunun, aktivite seviyeleri genetik veya çevresel değişimlere yanıt olarak değişen çeşitli geçici ya da kalıcı transkripsiyon faktörleri tarafından düzenlendiği bilinmektedir. Günümüzde değişimlere verilen yanıtların altında yatan karmaşık mekanizmayı keşfetmek büyük bir mücadeledir. Bu yüksek lisans tezinde, *Saccharomyces cerevisiae*' deki genetik ve çevresel değişimlere en anlamlı şekilde yanıt veren transkripsiyon faktörleri (etraflarında en anlamlı transkripsiyon değişimleri meydana gelen transkripsiyon faktörleri) varsayıma dayalı bir algoritma ile belirlenmiştir. Varolan yaklaşımlardan farklı olarak, bu tezde önerilen yaklaşım transkripsiyon faktörlerinin aktivitelerini belirlemede ağ yapısını kullanmaktadır. Bu çalışma çerçevesinde, *S. cerevisiae*' de çok kapsamlı bir transkripsiyonel düzenleyici ağ oluşturuldu ve gen ekspresyonu verisi ile bütünleştirildi. Maya verisinin analizi, kullanılan metodun değişimlere en anlamlı şekilde yanıt veren transkripsiyon faktörlerini belirlemede başarılı olduğunu göstermektedir. Buna ek olarak, bu metodla transkripsiyon faktörlerinin şartlara bağlı/geçici davranışları ile ilgili değerli bilgiler elde edilebilmiştir. Bu tez çalışmasında, değişimlere yanıt veren transkripsiyon faktörleri bulunduktan sonra değişimlere yanıt veren altağların da ortaya çıkarılabileceği gösterilmiştir. Altağlar değişimlere en anlamlı yanıt veren transkripsiyon faktörleri ve aynı değişimlere en anlamlı yanıt veren hedef genlerin birbirlerine bağlanmaları ile oluşturulmuştur. İlaveten, değişimlere en anlamlı yanıt veren her bir transkripsiyon faktörü için, aday hedef genleri öngörülmüştür (bu genler ve transkripsiyon faktörleri arasındaki etkileşimler henüz deneysel olarak kanıtlanmamıştır). Bu genler, transkripsiyon faktörleri ile aynı değişimlere yanıt vermiş olup aynı zamanda bu genlerde transkripsiyon faktörlerinin bağlanma yerleri bulunmaktadır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

C	Carbon
C	Constant
d	Diameter
k	Network connectivity
k	Number of connections of a transcription factor or a gene
k_{in}	In-degree
k_{out}	Out-degree
l	Path length
N	Nitrogen
p	Probability
P	Phosphorous
S	Sulfur
Z_{ni}	Z score of a gene node
Z_{TF}	Z score of a transcription factor node
β	In-degree exponential exponent
γ	Out-degree power-law exponent
Δ	Deletion
μ_k	Mean
σ_k	Standard deviation
AER	Aerobic
ANA	Anaerobic
CDF	Cumulative distribution function
DNA	Deoxyribonucleic acid
FC	Fold change
GO	Gene ontology
GRAM	Genetic regulatory modules
GRAS	Generally regarded as safe
GWLA	Genome-wide location analysis

mRNA	Messenger ribonucleic acid
qPCR	Quantitative real-time polymerase chain reaction
RNA	Ribonucleic acid
TF	Transcription factor
tRNA	Transfer ribonucleic acid
WT	Wild-type

1. INTRODUCTION

During the last century, the researchers have been provided with huge amount of knowledge about individual cellular components and their functions. However, despite this success, it is clear that usually a discrete biological function can not be attributed to an individual molecule. Instead, most biological characteristics arise from complex interactions between the cell's numerous constituents, such as proteins, deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and small molecules. Biology has now moved closer to the area of Systems Biology which seeks to integrate biological data as an attempt to understand how biological systems function. By studying the relationships and interactions between various parts of a biological system it is hoped that an understandable model of the whole system can be developed. Therefore in the twenty-first century, one of the main aims for biology is to understand the structure and the dynamics of the complex intracellular web of interactions that contribute to the structure and function of a living cell. In other words, network biology arises as a key branch in Systems Biology with the aim to understand the cell's functional organization. Rapid advances in network biology indicate that cellular networks are governed by universal laws and have the potential to revolutionize the view of biology and disease pathologies in the twenty-first century.

There are three types of intracellular biochemical reaction networks where reconstruction efforts are currently underway: metabolic, transcriptional regulatory and signaling networks. Ultimately, all three have to be integrated to generate whole-cell models of microbes and other organisms. The primary role of transcriptional regulation is the response to changes in environmental conditions, such as nutritional status and environmental stresses. Due to the central role that transcriptional regulation plays in cellular function and the availability of experimental techniques to elucidate regulatory networks, reconstruction of these networks has emerged as a key task in biology (Wyrick and Young, 2002).

Network analysis has recently been used in molecular biology, but so far almost all of the resultant networks have only been analyzed statically (Jeong *et al*., 2000; Guelzim *et al*., 2002; Milo *et al*., 2002; Shen-Orr *et al*., 2002; Oltvai *et al*., 2002; Barabasi *et al*., 2004;

Milo *et al*., 2004; Teichmann *et al*., 2004). The only study on the dynamics of the transcriptional regulatory networks revealed that over half of the active interactions between transcription factors (TFs) and their target genes are found to be completely replaced by new ones between conditions (Luscombe *et al*., 2004). This result shows that an analysis of the dynamics of a network is necessary considering significant changes occur in a network between different conditions.

Here, an integrative computational approach is reported in order to identify perturbation-responsive key transcription factors in transcriptional regulatory networks. The algorithm used has recently been developed and implemented succesfully for identification of so-called reporter metabolites (metabolites around which the most significant transcriptional changes occur) in yeast by the integration of the genome-scale metabolic network with transcriptome data (Patil and Nielsen, 2005). In the present study, the approach was modified so that perturbation-responsive key transcription factors could be identified using the same algorithm. In the modified approach, genome-wide transcriptome data is comparatively (two conditions or two strains are compared) integrated with the transcriptional regulatory network constructed. Previous studies on the dynamics of transcription factors focused on the activity of transcription factors under certain conditions and thus on identification of condition-specific transcription factors (Wang *et al*., 2002; Segal *et al*., 2003; Luscombe *et al*., 2004) but not on the effects of perturbations (*i.e.* passing from one condition to another) on the activity of transcription factors. Our approach enables the identification of transcription factors which respond significantly to a genetic or an environmental perturbation (*i.e.* TFs activated/deactivated when passing from one condition to another). Identifying these transcription factors is crucial for understanding the regulatory mechanisms in a cell subjected to a perturbation. Moreover, these transcription factors may relate to perturbation-dependent lethality and their identification may lead to determination of specific drug targets. In fact, the necessity and the importance of the identification of oxygen-responsive transcription factors has recently been reported (Tai *et al*., 2005). An advantage of our method over the previous methods for inferring regulatory networks from gene expression data (Dhaseieer *et al*., 2000; Pe'er *et al*., 2001; Hartemink *et al*., 2002; Tanay and Shamir, 2001; Segal *et al*., 2003; Luscombe *et al*., 2004), is that it does not rely on the assumption that the activity levels of the transcription factors can be determined directly from their expression profiles.

Clearly, this assumption is violated in case of transcription factors which are regulated post-translationally. Instead, in our approach, the activities of TFs are determined by the the changes that occur around them (*i.e.* changes in the expression levels of their target genes). Hence, in this approach, contrary to previous ones, network topology is used to determine the activity levels of TFs.

Saccharomyces cerevisiae is the most thoroughly investigated eukaryotic microorganism, which aids our understanding of the biology of the eukaryotic cell and hence, ultimately, human biology. *S. cerevisae* is a very attractive organism to work with since it is nonpathogenic, and due to its long history of application in the production of consumable products such as ethanol and baker's yeast, it has been classified as a GRAS organism (generally regarded as safe). Also, the well-established fermentation and process technology for large-scale production with *S. cerevisiae* makes this organism attractive for several biotechnological purposes. Another important reason for the applicability of *S. cerevisiae* within the field of biotechnology is its susceptibility to genetic modifications by recombinant DNA technology, which has been further facilitated by the availability of the complete genome sequence of *S. cerevisiae*, published in 1996 (Ostergaard *et al*., 2000).

Based on the arguments above, the aim of this study is to uncover and elucidate the transcriptional regulatory mechanism in yeast *S. cerevisiae* from a dynamic perspective by:

- Creating a genome-scale map of transcriptional regulatory network in *S. Cerevisiae* and analysing its topology.
- Identifying perturbation-responsive key transcription factors, which are TFs around which most significant transcriptional changes occur, by integrating the transcriptional regulatory network constructed with transcription data for diverse conditions. The transcription data are comparatively integrated so that effects of perturbations, namely environmental and genetic perturbations such as change in oxygen availability, change in macronutrient limitation and deletion of a gene enconding a TF, are reflected in the network.
- Identifying perturbation-responsive subnetworks in the genome-scale transcriptional regulatory network, by interconnecting the perturbation-responsive key TFs

previously determined and their target genes which are differentially expressed when the same perturbation is introduced.

- Identifying best candidate target genes for each key TF. There is a very large number of genes whose promoter sequences match with the binding site of the TF. These genes are so-called potential target genes of the TFs. The best candidate target genes are the genes which are upregulated or downregulated significantly when the same perturbation is introduced to the system and are selected among the potential target genes listed in the database YEASTRACT.

2. THEORETICAL BACKGROUND

2.1. An Overview of Gene Control and Transcription Factors

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control (Alberts *et al*., 1994).

Transcription factors (also called gene regulatory proteins) recognize short stretches of double-helical DNA of defined sequence and thereby determine which of the thousands of genes in a cell will be transcribed. Thousands of TFs have been identified in a wide variety of organisms. Current estimates suggest that there are just over 200 DNA-binding transcription factors in yeast (Eisenstein, 2005). Although each of these proteins has unique features, most bind to DNA as homodimers or heterodimers and recognize DNA through one of a small number of structural motifs. The common motifs include the helix-turn-helix, the homeodomain, the leucine zipper, the helix-loop-helix and zinc fingers of several types. The precise amino acid sequence that is folded into a motif determines the particular DNA sequence that is recognized. Heterodimerization increases the range of DNA sequences that can be recognized by TFs. Powerful techniques are available that make use of the DNA sequence specifity of TFs to identify and isolate these proteins, the genes that encode them, the DNA sequences they recognize and the genes that they regulate (Alberts *et al*., 1994).

The transcription of individual genes is switched on and off in cells by TFs. In procaryotes these proteins usually bind to specific DNA sequences close to the RNA polymerase start site and depending on the nature of the TF and the precise location of its binding site relative to the start site, either activate or repress transcription of the gene. The

flexibility of the DNA helix, however, also allows proteins bond at distant sites to affect the RNA polymerase at the promoter by the looping out of the intervening DNA. Such action at a distance is extremely common in eucaryotic cells, where TFs bound to sequences thousands of nucleotide pairs from the promoter generally control gene expression. Eucaryotic activators and repressors act by a wide variety of mechanisms – generally causing the local modification of chromatin structure, the assembly of the general TFs at the promoter and the recruitment of RNA polymerase.

Whereas the transcription of a typical procaryotic gene is controlled by only one or two TFs, the regulation of higher eucaryotic genes is much more complex with the larger genome size and the large variety of cell types that are formed. Some of the TFs are transcriptional activators, whereas others are transcriptional repressors. These proteins bind to regulatory sequences organized in a series of regulatory modules strung together along the DNA and tohether they cause the correct spatial and temporal pattern of gene expression. Eucaryotic genes and their control regions are often surrounded by insulators, DNA sequences recognized by proteins that prevent cross-talk between independently regulated genes (Alberts *et al*., 1994).

2.2. Transcriptional Regulatory Networks

Analysis of the changes in the level of gene expression (mRNA and protein) provide insights into regulatory influences and not necessarily mechanisms responsible for mediating those changes. To approach a mechanistic level it is imperative to map the physical linkages among proteins (protein-protein interactions) and proteins and DNA (protein-DNA interactions). The transcriptional regulatory network is then defined by which transcription factor binds to which promoters and what the integrated effect of all these transcription factors is on the expression of all the genes. The basic functional element of a regulatory network is the promoter region of a gene or operon, which contains the regulatory binding sites for the relevant transcription factors that regulate the expression of a particular gene. The locations and orientations of these binding sites, also the affinity of the transcription factors to particular variants of the site, determine the expression levels of a gene in response to changes in the active transcription factor concentrations inside the cell (Ptashne and Gann, 2002).

It has been demonstrated that the known organization of promotor regions in bacteria allows the implementation of a wide class of regulatory logic functions within a single promoter (Buchler *et al*., 2003), so that even a single 'node' in the regulatory network can be relatively complex. At the basic level the mechanisms of transcriptional regulation are the same for prokaryotes and eukaryotes, but eukaryotic organisms add an additional level of complexity to the regulatory network in the form of chromatin-modifying enzymes and other co-regulators that are typically recruited to promoters by specific transcription factors (Struhl, 1999).

For regulatory networks the number of TFs cannot be simply used to estimate the complexity of the network, owing to the fact that TFs can have multiple target genes and can often act in synergistic combinations (Herrgard *et al*., 2004). However, the relative fraction of TF coding genes tends to be higher for organisms that encounter more varied environmental conditions during their lifetime (Cases *et al*., 2003), indicating there are limits to the complexity that can be achieved with a fixed number of TFs. Information on well-studied organisms can be used to evaluate the level of complexity of transcriptional regulatory networks in terms of the number of components, transcription factors and target genes, and regulatory interactions. *Escherichia coli* has been predicted to have 314 transcription factors (Perez-Rueda and Collado-Vides, 2000) and on the basis of primary literature 1468 regulatory interactions have been identified (Shenn-Orr *et al*., 2002). In *S. cerevisiae,* there are 334 transcription factors (TRANSFAC, 2005) and large-scale *in vivo* protein-DNA binding screens indicate that there are at least 4000 regulatory interactions (Lee *et al*., 2002). For both *E. coli* and yeast these numbers are probably underestimates since significant efforts are still underway in this area but they give an indication of the order of magnitude of the regulatory network reconstruction task.

2.2.1. Databases and Experimental Data

For regulatory networks comprehensive databases covering genome-scale regulatory networks in multiple organisms do not currently exist (Herrgard *et al*., 2004). For individual organisms, however, such network databases containing experimentally verified regulatory interactions have been established, the most prominent one being RegulonDB for *E. coli* (Salgado *et al*., 2001). There are also general databases for individual

organisms, such as the Yeast Proteome Database (YPD) (Csank *et al*., 2002), that contain significant amounts of regulatory information, and YEASTRACT (Teixeira *et al*., 2006), very recently created repository of more than 12000 regulatory associations between transcription factors (TF) and target genes in *S. cerevisiae*. In addition to databases describing regulatory network structures, there are comprehensive databases that specialize in describing TF-binding sites, such as SCPD (Zhu and Zhang, 1999) for yeast and the general TF-binding database TRANSFAC (Matys *et al*., 2003). Although these databases contain valuable information for regulatory network reconstruction, they are not very complete and for the most part lack information about the synergistic effects between TFs acting on one gene. Nevertheless, these databases and primary information can be utilized to reconstruct regulatory networks for well-characterized organisms.

The major advantage that regulatory network reconstruction has over other types of network reconstructions, such as metabolic network reconstruction, is the availability of high-throughput experimental data that is directly relevant to the network structure. For metabolic processes the only widely available data source is the genome sequence and its annotation – techniques for measuring relevant metabolic quantities such as metabolic fluxes and metabolite levels are still not commonly used and have not been fully scaled to the whole-genome level (Stitt *et al*., 2003). By contrast, the two primary data types useful for the regulatory network reconstruction task – genome-wide mRNA expression and location analysis data – are widely available (Herrgard *et al*., 2004).

Gene expression data can be readily generated for well-studied microbial organisms using several standard technologies (Holloway *et al*., 2002). Advances in statistical data analysis allow both significant changes in gene expression under different conditions to be established (Quackenbush, 2002) and hypotheses about regulatory interactions or co-regulated gene modules to be derived directly from the data (Segal *et al*., 2003). Genome-wide location analysis (GWLA) is a method that allows protein-DNA interactions to be monitored across the entire yeast genome. The method combines a modified chromatin immunoprecipitation (ChIP) procedure, which has been previously used to study protein-DNA interactions at a small number of specific DNA sites, with DNA microarray analysis (Ren *et al*., 2000). GWLA, allowing the direct detection of genomic target sites for DNA-binding proteins such as TFs, promises to lead to an even more significant improvement in

our ability to reconstruct regulatory network structures than gene expression profiling. So far, GWLA has been most extensively applied in yeast, where it has been used to map the target genes of 106 TFs under one set of conditions (Lee *et al*., 2002). In principle the technique can be readily extended to other organisms (Laub *et al*., 2002). GWLA has also been used to study the stimulus-dependent binding of TFs (Zeitlinger *et al*., 2003), opening up the possibility of using this technique to map combinatorial interactions between TFs on a genome-wide scale.

The combination of expression profiling with GWLA as well as promoter sequence motif analysis has allowed the generation of hypothetical regulatory network structures using a variety of data integration methods (Lee *et al*., 2002; Segal *et al*., 2003; Hartemink *et al*., 2002; Liu *et al*., 2002). Deriving full regulatory network structures solely based on experimental data appears to be challenging, however, owing to the large quantities of high-quality data that would be required for such a task. One alternative to this purely data-driven approach would be to utilize well-curated regulatory network structures derived from databases and primary literature as a starting point for expanding the network on the basis of high-throughput data (Figure 2.1). For such an approach to succeed, one first needs to evaluate how well current known regulatory network structures agree with high-throughput datasets. This type of analysis has been performed for yeast and *E.coli* (Herrgard *et al*., 2003; Gutierrez-Rios *et al*., 2003; Yu *et al*., 2003). These studies have allowed the definition of network subcomponents and network structural motif types that are well supported by gene expression data and thus are good targets for data driven model expansion. In the future, such combinations of knowledge-driven and data-driven regulatory network reconstruction strategies may allow the acceleration of network reconstruction in well-studied organisms (Herrgard *et al*., 2004).

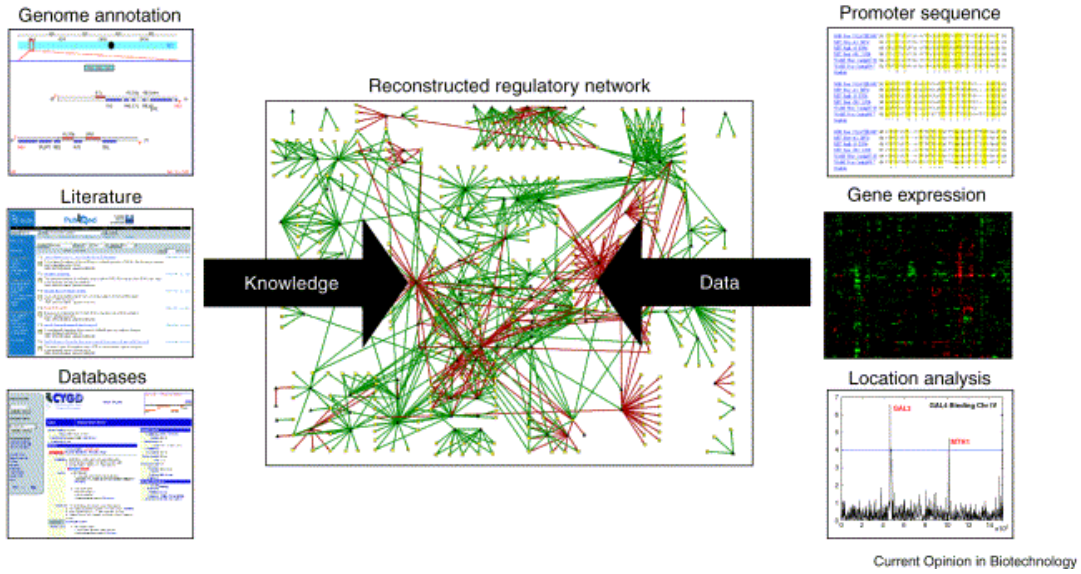

Figure 2.1. Knowledge and data based regulatory network reconstruction. Regulatory networks can be reconstructed by collecting individual regulatory interactions from relevant databases and the primary literature (knowledge). Alternatively, networks can be derived directly from high-throughput experimental data and promoter sequence analysis through various data-mining methods (Herrgard *et al*., 2004)

2.2.2. Module Networks: Identifying Regulatory Modules

The complex functions of a living cell are carried out through the concerted activity of many genes and TFs. This activity is often coordinated by the organization of the genome into regulatory modules, or sets of coregulated genes that share a common function. Such is the case for most of the metabolic pathways as well as for members of multiprotein complexes. Identifying this organization is crucial for understanding cellular responses to internal and external signals. Genome-wide expression profiles provide important information about these cellular processes. Yet, the regulatory machinery of the cell is far from transparent in these data. Current approaches for analyzing gene expression data allow the identification of groups of co-expressed genes. But the regulatory programs of these groups can be suggested only indirectly (Segal *et al*., 2003).

The first approach in module networks had the aim of reconstructing transcription modules (defined by a transcription factor and its target genes) and identifying conditions under which a particular transcription module is activated/deactivated (Wang *et al*., 2002).

The approach integrates information from regulatory sequences, genome-wide mRNA expression data and functional annotation. In this study, gene expression profiling experiments were systematically analyzed in which the yeast cell was subjected to various environmental or genetic perturbations. They were able to construct transcription modules with high specificity and sensitivity for many transcription factors and predict the activation of these modules under both anticipated and unexpected conditions. Correlating the activation of a module to a specific perturbation predicts links in the cell's regulatory networks.

Same year, an alternative method was proposed for the global analysis of genome-wide expression data, arguing that standard clustering methods can classify genes successfully when applied to small data sets but have limited use in the analysis of large-scale expression data (Ihmels *et al*. 2002). The approach assigns genes to context-dependent and potentially overlapping transcription modules (which are defined as combined groups of genes and conditions), thus overcoming the main limitations of traditional clustering methods. The method is based on an algorithm that receives a gene set that partially overlaps a transcription module and then provides the complete module as output. The algorithm is referred as the signature algorithm. The method was used to elucidate regulatory properties of cellular pathways and to characterize cis-regulatory elements. By applying the algorithm systematically to all of the available expression data on *S. cerevisiae,* a comprehensive set of overlapping transcriptional modules was identified. The results provided functional predictions for numerous genes, identified relations between modules and presented a global view on the transcriptional network.

One year later, a complementary method to the above mentioned method was proposed (Bergmann *et al*., 2003). The main idea of the previous work was to integrate prior biological information, such as the function or sequence of known genes, into the analysis of the gene expression data. However, the new method, called iterative signature algorithm, did not require any prior knowledge beyond the expression data.

In order to describe a genome-wide regulatory network in *S. cerevisiae* an algorithm for discovering regulatory networks of gene modules, so-called GRAM (Genetic Regulatory Modules) algorithm was developed (Bar-Joseph *et al*., 2003). This algorithm

combines information from genome-wide location and expression data sets. A gene module is defined as a set of co-expressed genes to which the same set of transcription factors binds. Unlike many approaches that rely primarily on functional information from expression data, the GRAM algorithm explicitly links genes to the factors that regulate them by incorporating DNA binding data, which provide direct physical evidence of regulatory interactions. The GRAM algorithm was used to describe a genome-wide regulatory network in *S. cerevisiae* using binding information for 106 transcription factors profiled in rich medium conditions using data from more than 500 expression experiments.

2.2.3. Network Motifs in Transcriptional Regulatory Networks

Network motifs are precise patterns of inter-connections between a small number of TFs and target genes (Shen-Orr *et al*., 2002). Three of the most common motifs are single input motif (SIM), multiple input motif (MIM) and feed-forward loop (FFL) (Figure 2.2). The network motifs were first investigated in the transcriptional regulation network of *E. coli* (Shen-Orr *et al*., 2002). In this study, network motifs were described as patterns of interconnections that recured in many different parts of a network at frequencies much higher than those found in randomized networks and in order to systematically detect network motifs to one of the best-characterized regulation networks, that of transcriptional interactions in *E. coli*, an algorithm was proposed. This study revealed that much of the network was composed of repeated appearances of three highly significant motifs. Each network motif had a specific function in determining gene expression, such as generating temporal expression programs and governing the responses to fluctuating external signals. On the other hand, the motif structure allowed an easily interpretable view of the entire known transcriptional network of the organism. It was found that SIMs and MIMs were implicated in conferring similar regulation over groups of genes, so they were ideal for directing the large-scale gene activation. FFLs were found to be buffers that responded only to persistent input signals (Shen-Orr *et al*., 2002).

The level of co-expression between the genes is very dependent on the type of regulatory network motif. Genes targeted by individual TFs (SIM) are not strongly correlated: just 1.3 per cent of target pairs are co-expressed although this is significantly higher than expected. Correlation is stronger for genes targeted by multiple, common TFs:

24.4 per cent of MIM target pairs and 5.0 per cent of FFL targets ehibit co-expression. Expression is much more tightly regulated when multiple TFs are involved. Ndd1p network is an example for SIM type TF-target relationship. Ndd1p, a cell cycle regulator, acts as the sole regulating TF for Mcm21p, a kinetochore protein required for normal cell growth and *STB5*, encoding another transcription factor. Forkhead network is an example for MIM type TF-target relationship. Ndd1p is recruited to promoters by Fkh1p and Fkh2p, two forkhead transcription activators. Collectively, these three TFs regulate Dbf2p, a kinase needed for cell cycle regulation, and HDR1 (function unknown). Finally, Mbp1p/Swi4p network can be given as an example for FFL type TF-target relationship. In a feed-forward-loop, Mbp1p (a cell cycle regulator controlling DNA replication and repair) is the leading TF, Swi4p (a cell cycle regulator controlling cell wall and membrane synthesis) is the intermediate TF, and *SPT21* (a TF involved in histone expression) and *YML102C-A* (function unknown) are the target genes (Yu *et al*., 2003).

Recently, it has been reported that the relative occurrence of motifs varied considerably between different conditions (Luscombe *et al*., 2004). SIMs were found to be favoured in subnetworks active during diauxic shift, DNA damage and stress response, where they comprised >55 per cent of regulatory interactions in motifs. But the frequency droped to 35 per cent during cell cycle and sporulation. Instead, these states favoured FFLs (44 per cent). It is notable that MIMs did not significantly change their usage (Luscombe *et al*., 2004).

network motif	abbreviation	schematic
single input motif	*SIM*	
multiple input motif	*MIM*	
feed-forward loop	*FFL*	

Figure 2.2. Most common motifs in regulatory networks (Luscombe *et al*., 2004)

2.2.4. From Network Reconstructions to Mathematical Models

The first class of modeling approaches are primarily intended to describe the structural features of regulatory networks and do not accurately predict gene expression levels in response to changes in regulatory activity. Directed graphs with transcription factors and target genes as nodes and regulatory interactions as edges are commonly used to visualize regulatory netwoks and to analyze their structural properties (Shen-Orr *et al*., 2002; Guelzim *et al*., 2002). Most methods for reconstructing regulatory networks based on gene expression and/or GWLA data describe the regulatory network as a directed graph (Lee *et al*., 2002). These graphs cannot represent important interactions between transcription factors and they do not allow simulation of model behavior or effective integration of regulatory networks with models of other cellular processes. However, the graph-based models of regulatory networks can also be used as a basis for building more quantitative models through measuring the regulatory strengths for different regulatory interactions experimentally (Ronen *et al.*, 2002).

The second class of modeling approaches focuses on the prediction of gene expression levels at the expense of the scale of regulatory network subcomponents that can be modeled. Linear differential equations or linear models relating transcription factor and

target expression levels are the simplest of these approaches. This type of linear model was utilized in a recent study of the SOS response system in *E. coli*, in which experimental gene expression data was used to directly reconstruct a model for a small regulatory network without any prior knowledge of the network structure (Gardner *et al*., 2003). Systematic transcriptional perturbations were introduced to construct a first-order model of regulatory interactions in a nine-gene subnetwork of the SOS pathway. Quantitative real-time polymerase chain reaction (qPCR) was used to measure the change in expression of perturbed cells relative to unperturbed cells. The constructed model was successful in predictions. This approach provided a framework for elucidating the functional properties of genetic networks and identifying molecular targets of pharmacological compounds.

As the role of transcriptional regulation is to modulate other cellular processes, integrating the reconstructed regulatory networks with models of these other processes is central to understanding regulatory network function in the context of all organism. The major advantage with such integrated models is that even when the modeling of the regulatory network is done at the qualitative level, the integrated regulatory/metabolic model can be used to quantitatively predict phenotypes such as growth rates. Furthermore, comparisons between model predictions and experimental data can be used to improve the model systematically. These types of integrated model are a powerful way to bring together multiple types of high-throughput data (e.g. gene expression and phenotyping) and to interpret these datasets, as discrepancies between model predictions and experimental data can point to specific inconsistencies in the current reconstructed regulatory network model (Herrgard *et al.*, 2004).

3. METHODS

The approach used to identify key transcription factors is a modified version of the approach recently developed for identification of reporter metabolites in yeast (Patil and Nielsen, 2005). The reporter metabolites were identified by integrating the metabolic network with gene-expression data. In this sudy, the approach was modified so that the transcriptional regulatory network constructed was integrated with transcription data and perturbation-responsive key transcription factors were identified. The modified approach for identification of key TFs consists of the following steps.

3.1. Construction of the Transcriptional Regulatory Network

A genome-scale transcriptional regulatory network in *S. cerevisiae* was constructed by assembling known regulatory interactions from several data sources (Lee *et al*., 2002; Wingender *et al*., 2001; Matys *et al*., 2003; Zhu and Zhang, 1999; YEASTRACT, 2005). The regulatory interactions are between TFs and non-TF targets or two TFs. Therefore, the interactions were represented as two columns, one representing the TFs and the other one representing the genes. Then, the complete transcriptional regulatory network was represented as a bipartite undirected graph. In this graph, both TFs and genes are represented as nodes and interactions between them as edges. In other words, a TF node is connected to all of the genes that it regulates, and a gene node is connected to all of the TFs that are known to regulate this gene.

The topological measures of the network, such as *in-degree* ($<k_{in}>$), *out-degree* ($<k_{out}>$), *in-degree exponential exponent* (β), *out-degree power-law exponent* (γ), *path length (l)* and *diameter (d),* were calculated. A computer code was written in FORTRAN in order to perform the calculations of the *in-degree* ($<k_{in}>$), the *out-degree* ($<k_{out}>$), the *path length (l)* and the *diameter (d)* (Appendix A). *In-degree* ($<k_{in}>$) is the number of incoming edges per node (*i.e.* the number of TFs regulating a target) and *out-degree* ($<k_{out}>$) represents the number of outgoing edges per node (*i.e.* the number of target genes for each TF). The most suitable distributions for the in- and out-degrees were found by plotting P_k

values versus $<k_{in}>$ and $(<k_{out}>)$ values, where P_k is the probability that a randomly picked node has k interactions and is calculated by dividing the number of nodes which have k connections by the total number of nodes in the network. The best fitting distribution was found by fitting the plots of P_k to both exponential and power-law distributions. The fitted distribution which best represents the actual distribution (according to R-squared values of the plots) was chosen. The connectivity of the network was calculated using the following formula,

$$k = \frac{2l}{n} \tag{3.1}$$

where l is the number of edges and n is the number of nodes in the network.

The *path length (l)* is the shortest distance (in number of intermediate nodes) between two nodes and the *diameter (d)* is the maximum path length in the network.

3.2. Mapping and Scoring of Transcription Data

Differential transcription data in triplicates were used, in which two different strains or conditions are compared with three measurements for each strain or condition (Tai *et al*., 2005). Two types of perturbations, namely genetic and environmental perturbations, were analysed. In order to identify key TFs responsive to a genetic perturbation, the *MIG1* gene deleted strain and both *MIG1* and *MIG2* genes deleted strain were compaired with the wild-type strain (original parent strain). For identification of key TFs responsive to environmental perturbations, anaerobic and aerobic conditions under four macronutrient limitation regimes (carbon, nitrogen, phosphorous and sulfur) were compared. In addition, the four macronutrient limitation regimes were pairwise compared for both aerobic and anaerobic conditions.

Each gene node of the transcriptional regulatory network was scored based on the probability that the differential expression is due to chance. Student's t test was used to obtain p values. $(1 - p_i)$ is the probability that the differential gene expression is not due to

chance. Then probabilities are converted to Z scores for each gene node (Z_{ni}) by using the inverse normal cumulative distribution (θ^{-1}).

$$Z_{ni} = \theta^{-1}(1 - p_i) \tag{3.2}$$

3.3. Method for Identification of Key Transcription Factors

For the identification of the key transcription factors, each TF node was scored based on the normalized transcriptional response of its neighboring genes (*i.e.* *Z* scores of the genes connected to the TF are added and the sum was multiplied by the inverse square root of the number of neighbors that each TF has). A computer code was written in FORTRAN in order to perform the calculations of *Z* scores (Appendix B).

$$Z_{TF} = \frac{1}{\sqrt{k}} \sum Z_{ni} \tag{3.3}$$

Z_{TF} scores were then corrected for the background distribution. 10,000 sets of *k* genes were randomly selected from the graph and new random Z_{TF} scores were calculated for each TF. Then, the mean (μ_k) of the resultant Z_{TF} scores were substracted from the previously calculated score and divided by their standard deviation (σ_k). A computer code was written in FORTRAN in order to perform the calculations of the corrected *Z* scores (Appendix C).

$$Z_{\underset{TF}{corrected}} = \frac{(Z_{TF} - \mu_k)}{\sigma_k} \tag{3.4}$$

Due to the correction, TFs which have a very large number of connections in the network and could therefore be identified as key TFs by mistake, regardless of the conditions, were eliminated. However, due to the large number of the connections that these TFs make, it is possible that they may be the hubs in the network (*i.e.* they may represent the most important points for the structure of the network) and hence their removal may significantly affect the network topology. The top 10 highest scoring TFs

before the correction are listed in Appendix D. After correcting the Z scores for background distribution, top 10 highest scoring TFs are identified as key TFs having great importance since most significant transcriptional changes occur around these TFs when a perturbation is introduced. The Gene Ontology (GO) terms associated with each key TF were found from the database YEASTRACT.

4. RESULTS AND DISCUSSION

4.1. Transcriptional Regulatory Network

The constructed genome-scale transcriptional regulatory network in *S. Cerevisiae* contained 9075 regulatory interactions between 180 TFs and 3514 targets. Interactions are between TFs and non-TF targets or two TFs. The network was visualized using the program Cytoscape 2.0 (Figure 4.1). Until now, relatively smaller transcriptional regulatory networks (*i.e.* containing less number of TFs and/or interactions) have been reported in the literature (Lee *et al*., 2002; Yu *et al*., 2003; Luscombe *et al*., 2004).

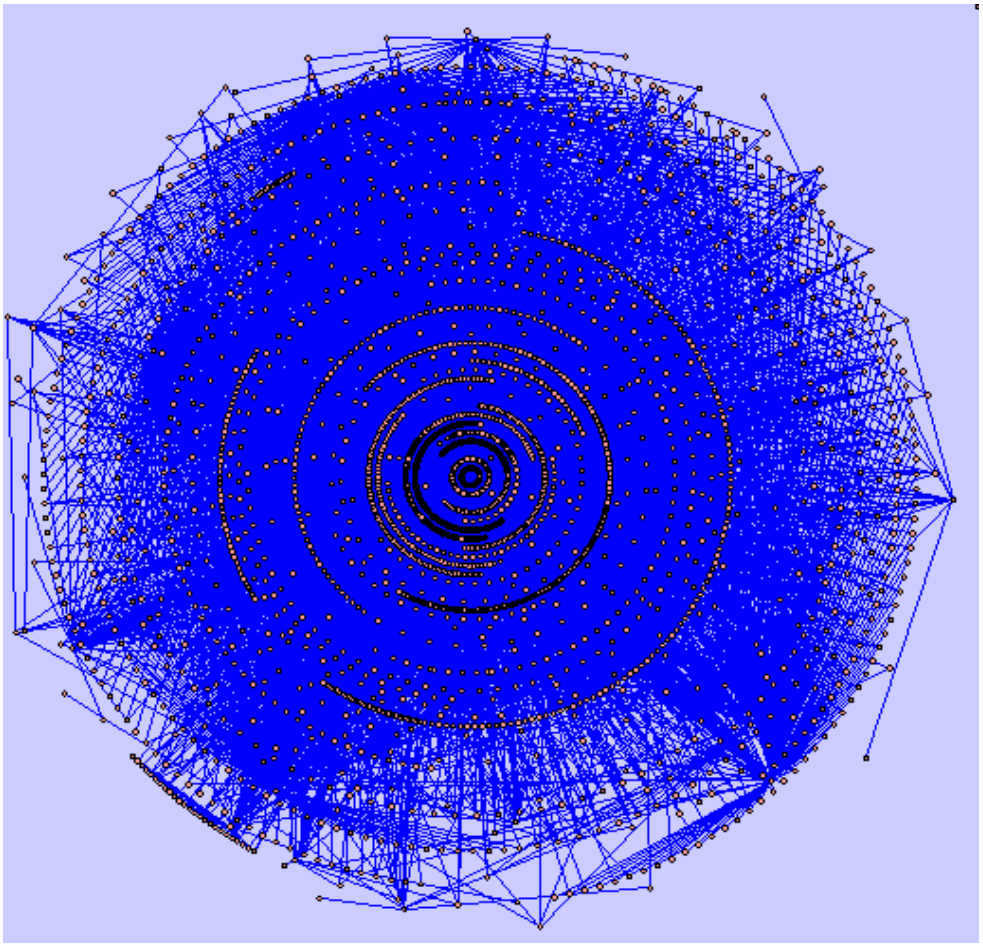

Figure 4.1. Constructed transcriptional regulatory network in *S. Cerevisiae*

The network constructed here is highly connected having a *connectivity (k)* of 4.91. The average of the *in-degree (k_{in})* (the number of TFs regulating a target) and the average of the *out-degree (k_{out})* (the number of target genes for each TF) are both equal to 2.43. Like many large-scale networks, it displays scale-free characteristics (the connectivity of the network is distributed according to a power law $p_k=Ck^{-\gamma}$, where *C* is a constant and γ=1.63) (Figure 4.2). The arriving connectivity of the network has an exponential distribution ($p_k=Ce^{-\beta k}$ where β=0.40) (Figure 4.3) whereas the departing connectivity better fits to a power law distribution ($p_k=Ck^{-\gamma}$ where γ=0.64) (Figure 4.4). The exponential behavior shows the molecular constraints on the number of TFs that can co-regulate at the same promoter, whereas the power-law behavior indicates a hub-containing network structure (Luscombe *et al*, 2004). In an earlier study on the structure of a much smaller transcriptional regulatory network in yeast, β and γ were found to be 0.45 and 1, respectively, and it was argued that lower β coefficients were predicted for organisms with more sophisticated genetic regulatory machinery (Guelzim *et al*., 2002). In addition, in a recent study on the dynamics of transcriptional regulatory networks, the exponent γ and the exponent β of the genome-scale trancriptional regulatory network of *S. cerevisiae* were found to be 0.6 and 0.8, respectively (Luscombe *et al*, 2004). Considering that a larger network containing more TFs and interactions was analysed in the present study, it is logical that both the exponent β and γ were found to be smaller than in the above studies, and hence, the genetic regulatory machinery analysed in this study may be accepted as more sophisticated.

The *path length (l)* and the *diameter (d)* of the network were found to be 4.25 and 13, respectively. The diameter was found larger than the diameter of the most recently reported transcriptional regulatory network in literature (Luscombe *et al*., 2004), showing once more the largeness and complexity of the network constructed.

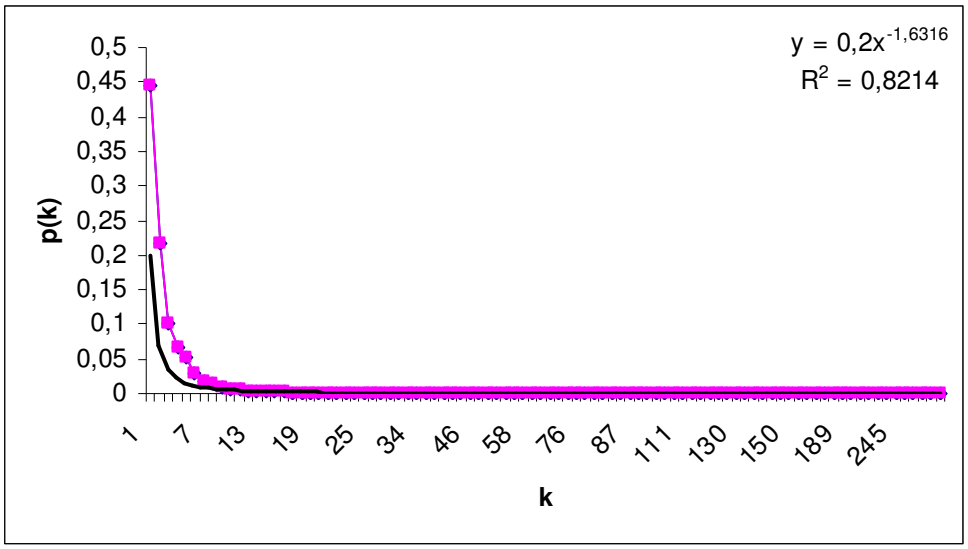

Figure 4.2. Power-law behavior of the network connectivity

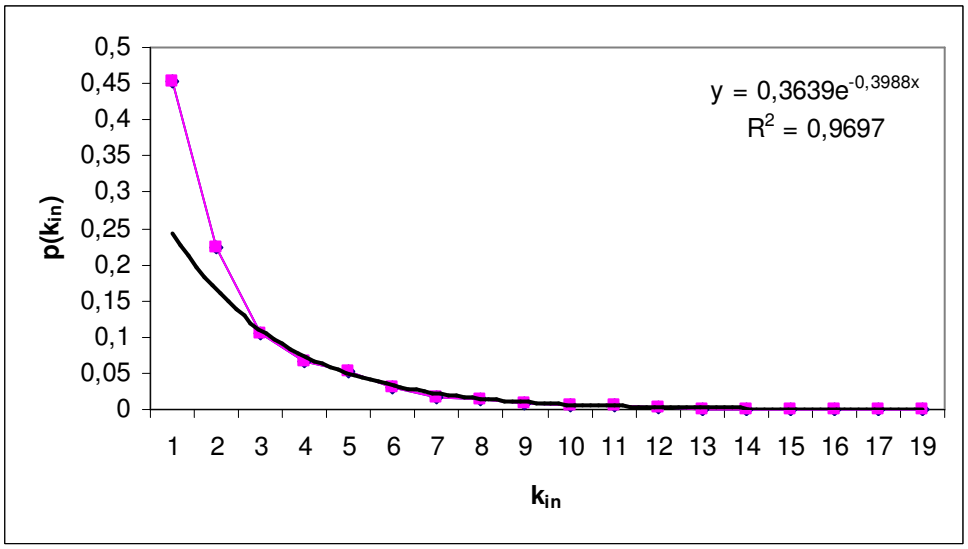

Figure 4.3. Exponential behavior of the network arriving connectivity

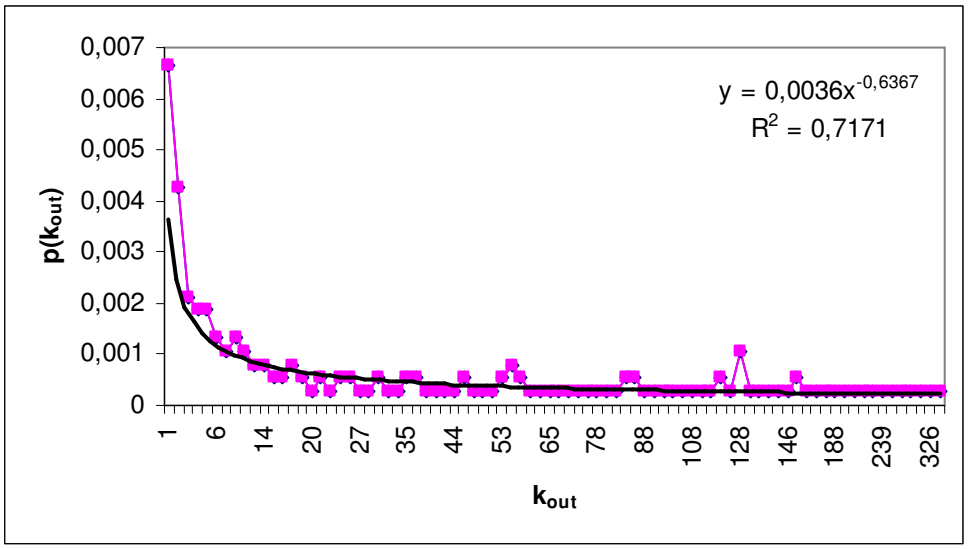

Figure 4.4. Power-law behavior of the network departing connectivity

4.2. Key Transcription Factors

4.2.1. Key TFs Responsive to Oxygen Availability

Firstly, the key TFs responding to change in the oxygen availability under a specified macronutrient limitation and the corresponding GO terms were identified (Tables 4.1 - 4.4). These TFs were the top 10 highest scoring TFs when shifting from anerobic condition to aerobic condition. Hap2p, Hap3p and Hap4p and Hap5p were counted as one transcription factor because they form a complex and their scores are very close to each other. Transcription data in triplicates for anaerobic and aerobic conditions for carbon, nitrogen, phosphorous and sulfur limitations were used (Tai *et al*., 2005).

Three TFs, namely Hap2/3/4/5p, Upc2p and Yap1p, were identified as key TFs between anaerobic and aerobic conditions under all four macronutrient limitations, showing a consistent response to oxygen availability independent of the nutrient limitation (Figure 4.5). This result is biologically meaningful since Hap2/3/4/5p complex is known to be involved in the fermentation-respiration shift (Bourgarel, 1999), Upc2p is known to regulate the majority of anaerobically induced genes in *S. cerevisiae* (Kwast, 2002), and recent experimental studies indicated involvement of Yap1p in response to oxidative stress (Rodrigues-Pousada, 2004).

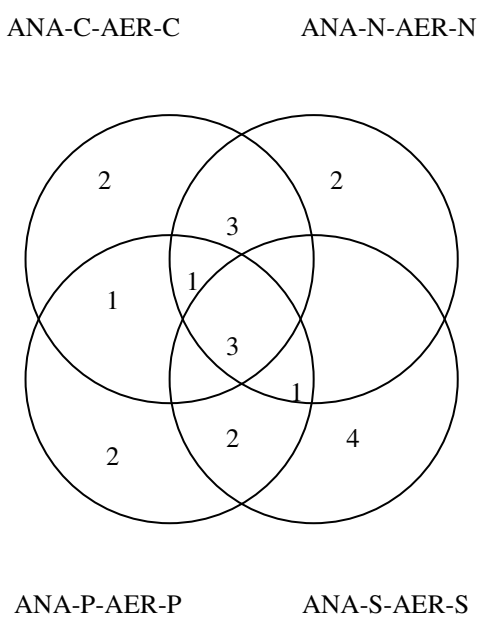

Figure 4.5. Venn diagram of oxygen-responsive TFs. Each of the four circles represents a cluster of TFs that were identified as key TFs under carbon, nitrogen, phosphorous or sulphur limitation regimes

Key transcription factors specific to one macronutrient limitation regime between anaerobic and aerobic conditions (*i.e.* showing response to oxygen availability under that specified nutrient limitation) also yielded meaningful information consistent with literature. When passing from anaerobic condition to aerobic condition under carbon limitation regime, Mig2p, Cat8p, Msn2p, Msn4p were among the key TFs identified (Table 4.1). These results are logical since Mig2p is involved in glucose metabolism, Cat8p in gluconeogenesis (which is the biosynthesis of new glucose not from glycogen but from other metabolites) and both Msn2p and Msn4p are involved in response to oxidative stres (Table 4.1). The increased activity of Cat8p under the carbon limitation regime may indicate the start of gluconeogenesis in the cell where carbon source is limited.

When passing from anaerobic condition to aerobic condition under nitrogen, phosphorous and sulfur limitation regimes, Rox1p was identified as a key TF (Tables 4.2-4.4). This result is consistent with the recently reported role of Rox1p in regulating the majority of anaerobically induced genes in *S. cerevisiae* along with Upc2p (Kwast, 2002).

When passing from anaerobic condition to aerobic condition under the sulfur limitation regime, Hap1p was identified as a key TF. The GO term associated with this TF is aerobic respiration (Table 4.4), and it also has a major role in the repair of oxidative DNA damage (Barzilay *et al*., 1996), which explains the reason why Hap1p was identified as a key TF between anaerobic and aerobic conditions.

Table 4.1. Key TFs between anaerobic and aerobic conditions-carbon limitation

Z-score	TF	GO Terms: Biological Process
8,55	Cat8p	• gluconeogenesis • positive regulation of transcription from Pol II promoter
7,55	Oaf1p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
7,39	Pip2p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
6,72	Hap2p	• regulation of carbohydrate metabolism • transcription
6,69	Hap3p	• regulation of carbohydrate metabolism • transcription
6,66	Hap5p	• regulation of carbohydrate metabolism • transcription
6,60	Hap4p	• regulation of carbohydrate metabolism • transcription
6,50	Msn4p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
5,64	Msn2p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
5,49	Upc2p	• steroid metabolism • sterol biosynthesis
5,47	Yap1p	• transcription • response to oxidative stress
5,02	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter
4,84	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter

Table 4.2. Key TFs between anaerobic and aerobic conditions-nitrogen limitation

Z-score	TF	GO Terms: Biological Process
12,39	Upc2p	• steroid metabolism • sterol biosynthesis
10,04	Hap2p	• regulation of carbohydrate metabolism • transcription
10,00	Hap5p	• regulation of carbohydrate metabolism • transcription
9,89	Hap4p	• regulation of carbohydrate metabolism • transcription
9,85	Hap3p	• regulation of carbohydrate metabolism • transcription
6,08	Rox1p	• negative regulation of transcription from Pol II promoter
5,55	Yap1p	• transcription • response to oxidative stress • response to drug
4,99	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
4,91	Oaf1p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
4,75	Pip2p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
4,31	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter
4,13	Sok2p	• pseudohyphal growth
3,90	Stp2p	• positive regulation of transcription from Pol II promoter

Table 4.3. Key TFs between anaerobic and aerobic conditions-phosphorous limitation

Z-score	TF	GO Terms: Biological Process
9,43	Hap2p	• regulation of carbohydrate metabolism • transcription
9,37	Hap3p	• regulation of carbohydrate metabolism • transcription
9,26	Hap4p	• regulation of carbohydrate metabolism • transcription
9,25	Hap5p	• regulation of carbohydrate metabolism • transcription
7,73	Upc2p	• steroid metabolism • sterol biosynthesis
5,10	Rox1p	• negative regulation of transcription from Pol II promoter
4,56	Cat8p	• gluconeogenesis • positive regulation of transcription from Pol II promoter
4,42	Met32p	• sulfur amino acid metabolism
4,35	Mot3p	• transcription
3,98	Ino2p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
3,89	Yap1p	• transcription • response to oxidative stress • response to drug
3,67	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter
3,56	Rgt1p	• glucose metabolism • regulation of glucose import

Table 4.4. Key TFs between anaerobic and aerobic conditions-sulfur limitation

Z-score	TF	GO Terms: Biological Process
10,04	Upc2p	• steroid metabolism • sterol biosynthesis
9,16	Hap5p	• regulation of carbohydrate metabolism • transcription
9,02	Hap2p	• regulation of carbohydrate metabolism • transcription
8,81	Hap3p	• regulation of carbohydrate metabolism • transcription
8,76	Hap4p	• regulation of carbohydrate metabolism • transcription
5,64	Yap1p	• transcription • response to oxidative stress • response to drug
5,06	Rox1p	• negative regulation of transcription from Pol II promoter
4,32	Mot3p	• transcription
4,05	Ino2p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
4,01	Rtg3p	• transcription initiation from Pol II promoter
3,94	Pdr3p	• regulation of transcription from Pol II promoter • response to drug
3,88	Hap1p	• aerobic respiration • positive regulation of transcription from Pol II promoter
3,48	Cad1p	• positive regulation of transcription from Pol II promoter • response to cadmium ion

4.2.2. Key TFs Responsive to Nutrient Limitation-Aerobic

Secondly, transcription data obtained in triplicates under aerobic conditions for carbon, nitrogen, phosphorous and sulphur limitations (Tai *et al*., 2005) were comparatively analysed. The transcription factors which were identified as key TFs when passing from one nutrient limitation to another under aerobic conditions are listed in Tables 4.5 - 4.10. Hap2p, Hap3p and Hap4p and Hap5p were counted as one transcription factor because they form a complex and their scores are very close to each other.

Only Mig2p was identified as the common key TF when passing from carbon limitation regime to any of the other three nutrient limitation regimes under aerobic

conditions (Figure 4.6a). This result is consistent with the known involvement of Mig2p in glucose metabolism (Table 4.5), and Mig2p is required for repression of many glucose-repressed genes (Lutfiyya *et al*., 1998).

Two TFs, namely Gln3p and Dal80p, were identified as the common key TFs when comparing the nitrogen limitation regime against the three others (Figure 4.6b). This prediction is supported by the fact that both Gln3p and Dal80p are involved in the regulation of the nitrogen utilization process (Chan *et al*., 2001; Svetlov and Cooper, 1998), besides it is remarkable that these two TFs, involved in nitrogen utilization process, were identified as key TFs only when comparing nitrogen limitation regime with others.

Stp1p and Stp2p were suggested as the common key TFs between phosphorous limitation regime and any of the three other limitation regimes (Figure 4.6c). These two TFs do not have a known relation to the phosphorous metabolism; however, they are both known to play a role in the positive regulation of transcription from Pol II promoter and were identified expectedly as key in most of the comparative sets.

When passing from the sulfur limitation regime to any of the other three limitation regimes, Gcn4p, Met32p, Pdr3p and Yap1p were identified as the common four TFs, or in other words, were suggested as responsive TFs to sulfur limitation regime (Figure 4.6d). Indeed, Met32p is involved in yeast sulfur amino acid metabolism (Blaiseau *et al*, 1997) and Gcn4p was recently identified as a nutrient deprivation responsive TF (Harbison *et al*., 2004). Yap1p is involved in the oxidative stress tolerance, and Pdr3p, being involved in positive regulation of transcription from Pol II promoter, was identified as key TF in most of the pairwise comparisons.

MATalpha2, which was recently suggested to take role in respiraton and carbon regulation (Segal *et al*., 2003), was identified as the key TF between carbon limitation and nitrogen limitation regimes under aerobic conditions. MATalpha2 makes only two connections in the network (*i.e.* it has two target genes). The fact that it was identified as a key TF shows the ability of the approach proposed in this work in capturing key TFs which make very few connections in the regulatory network. On the other hand, TFs with very

large number of connections in the network, such as Rap1p, which makes 483 connections, were not identified as key TFs since they were not active within the conditions analysed.

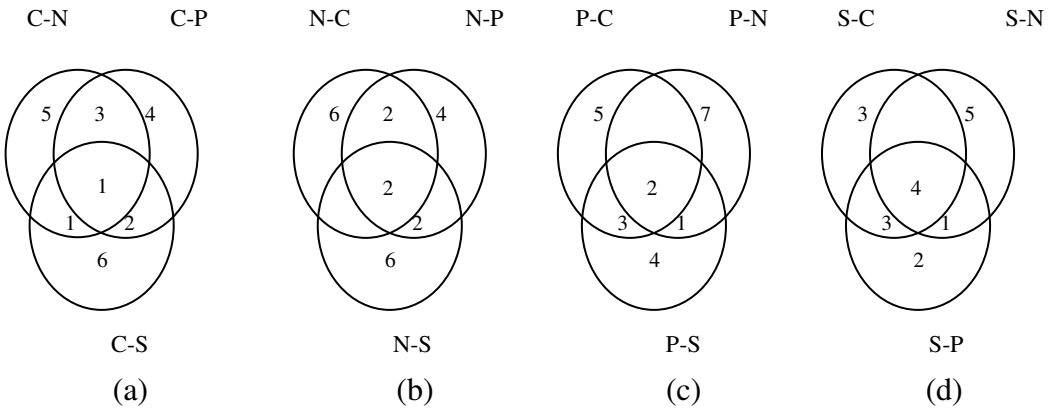

Figure 4.6. Venn diagrams of nutrient-responsive TFs (aerobic). Pairwise comparisons of each macronutrient (carbon, nitrogen, phosphorous and sulfur) limitation regime against the three others under aerobic conditions. Numbers indicate the number of key TFs.

Table 4.5. Key TFs between carbon and nitrogen limitations-aerobic

Z-score	TF	GO Terms: Biological Process
7,04	Gln3p	• nitrogen compound metabolism • regulation of nitrogen utilization • positive regulation of transcription from Pol II promoter
6,58	Dal80p	• transcription • regulation of nitrogen utilization
5,67	Pip2p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
5,15	Oaf1p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
4,94	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter
4,52	Cat8p	• gluconeogenesis • positive regulation of transcription from Pol II promoter
4,34	Gat1p	• transcription initiation from Pol II promoter • regulation of nitrogen utilization
4,19	Mcm1p	• DNA replication initiation • regulation of transcription from Pol II promoter
4,02	Stp2p	• positive regulation of transcription from Pol II promoter
3,95	MAT alpha2	• Homeobox-domain containing protein which, in haploid cells, acts with Mcm1p to repress a-specific genes

Table 4.6. Key TFs between carbon and phosphorous limitations-aerobic

Z-score	TF	GO Terms: Biological Process
7,51	Oaf1p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
7,00	Pip2p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription
6,89	Msn4p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
6,13	Stp2p	• positive regulation of transcription from Pol II promoter
6,02	Pdr3p	• regulation of transcription from Pol II promoter • response to drug
5,86	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
5,55	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter
5,23	Msn2p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
5,00	Ino2p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
4,96	Rtg1p	• transcription initiation from Pol II promoter • protein localization

Table 4.7. Key TFs between carbon and sulfur limitations-aerobic

Z-score	TF	GO Terms: Biological Process
5,67	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
5,63	Cat8p	• gluconeogenesis • positive regulation of transcription from Pol II promoter
5,24	Yap1p	• transcription • response to oxidative stress • response to drug
5,18	Pdr3p	• regulation of transcription from Pol II promoter • response to drug
5,03	Met32p	• sulfur amino acid metabolism
5,00	Met4p	• sulfur amino acid metabolism • positive regulation of transcription from Pol II promoter
5,00	Met28p	• sulfur amino acid biosynthesis • regulation of transcription from Pol II promoter • regulation of sulfur metabolism
4,96	Msn4p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
4,79	Cbf1p	• DNA replication and chromosome cycle • methionine biosynthesis
4,78	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter

Table 4.8. Key TFs between nitrogen and phosphorous limitations-aerobic

Z-score	TF	GO Terms: Biological Process
7,70	Gln3p	• nitrogen compound metabolism • regulation of nitrogen utilization • positive regulation of transcription from Pol II promoter
7,59	Dal80p	• transcription • regulation of nitrogen utilization
6,56	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
5,52	Stp2p	• positive regulation of transcription from Pol II promoter
4,98	Ste12p	• conjugation with cellular fusion • invasive growth (sensu Saccharomyces) • pseudohyphal growth • positive regulation of transcription from Pol II promoter by pheromones
4,78	Hap1p	• aerobic respiration • positive regulation of transcription from Pol II promoter
4,44	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
4,22	Gat1p	• transcription initiation from Pol II promoter • regulation of nitrogen utilization
4,05	Rtg3p	• transcription initiation from Pol II promoter
4,01	Pho4p	• phosphate metabolism • cellular response to phosphate starvation

Table 4.9. Key TFs between nitrogen and sulfur limitations-aerobic

Z-score	TF	GO Terms: Biological Process
7,56	Gln3p	• nitrogen compound metabolism • regulation of nitrogen utilization • positive regulation of transcription from Pol II promoter
7,12	Yap1p	• transcription • response to oxidative stress • response to drug
6,90	Pdr3p	• regulation of transcription from Pol II promoter • response to drug
6,82	Ino2p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
6,75	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
6,43	Dal80p	• transcription • regulation of nitrogen utilization
6,20	Ino4p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
6,00	Dal81p	• regulation of transcription from Pol II promoter • nitrogen utilization
5,79	Met32p	• sulfur amino acid metabolism
5,65	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter

Table 4.10. Key TFs between phosphorous and sulfur limitations-aerobic

Z-score	TF	GO Terms: Biological Process
7,70	Msn4p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
6,46	Met32p	• sulfur amino acid metabolism
6,33	Pdr3p	• regulation of transcription from Pol II promoter • response to drug
6,11	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
5,97	Yap1p	• transcription • response to oxidative stress • response to drug
5,75	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
5,57	Met28p	• sulfur amino acid biosynthesis • regulation of transcription from Pol II promoter • regulation of sulfur metabolism
5,57	Met4p	• sulfur amino acid metabolism • positive regulation of transcription from Pol II promoter
5,39	Stp2p	• positive regulation of transcription from Pol II promoter
4,71	Msn2p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress

4.2.3. Key TFs Responsive to Nutrient Limitation-Anaerobic

The transcription data obtained in triplicates for the four macronutrient limitation regimes, namely carbon, nitrogen, phosphorous and sulfur limitation regimes (Tai *et al*., 2005), were also comparatively analysed under anaerobic conditions. The TFs which were identified as key TFs when passing from one nutrient limitation to another under anaerobic conditions are listed in Tables 4.11 - 4.16. Hap2p, Hap3p and Hap4p and Hap5p were counted as one TF because they form a complex and their scores are very close to each other.

The TFs identified as common key TFs when comparing the carbon limitation regime and any of the other three limitation regimes were Hap2/3/4/5p and Rtg3p (Figure 4.7a). Hap complex is known to be involved in carbohydrate metabolism however, interestingly, it was not identified as key TF complex when comparing the nutrient limitation regimes under aerobic conditions. Rtg3p was, on the other hand, recently identified as a nutrient deprivation responsive TF (Harbison *et al*., 2004).

The three common key TFs between the nitrogen limitation regime and the three others were Dal80p, Dal81p, Gln3p and Stp1 (Figure 4.7b). This result is consistent with the fact that Dal80p, Dal81p and Gln3p are all involved in the regulation of the nitrogen utilization process (Svetlov and Cooper, 1998; Bricmont *et al*., 1991; Ogawa and Oshima, 1990). Besides, it is remarkable that these TFs involved in the regulation of the nitrogen utilization process were identified as key TFs only when comparing nitrogen limitation regime with others. As for Stp1p, as mentioned above, it is active in most of the pairwise comparisons being involved in positive regulation of transcription from Pol II promoter.

Stp1p and Stp2p were identified as common key TFs between phosporous limitation regime and the others, just as under aerobic conditions (Figure 4.7c). Gcn4p was also identified as key TF when passing from sulfur limitation regime to any other limitation regime (Figure 4.7d), as under aerobic conditions.

It is notable that Met4p, Met31p and Met32p, the three TFs involved in sulfur amino acid metabolism, were identified as key TFs only when passing from sulfur limitation regime to nitrogen and phosphorous limitation regimes (Tables 4.15 and 4.16). Moreover, Pho4p, a major TF in phosphate metabolism, was identified only between phosphorous limitation and sulfur limitation regimes (Table 4.16). These results show the ability of the present method to correctly identify the key TFs specific to conditions compared in this study.

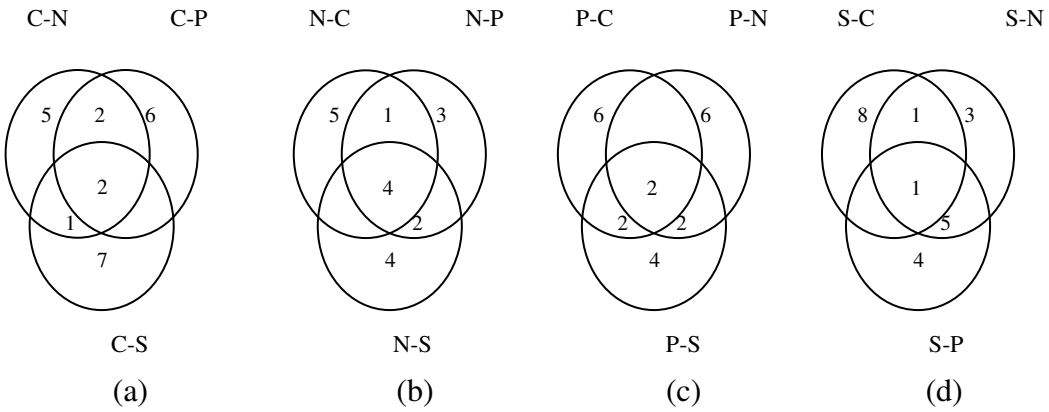

Figure 4.7. Venn diagrams of nutrient responsive TFs (anaerobic). Pairwise comparisons of each macronutrient (carbon, nitrogen, phosphorous and sulfur) limitation regime against the three others under anaerobic conditions. Numbers indicate the number of key TFs

Table 4.11. Key TFs between carbon and nitrogen limitations-anaerobic

Z-score	TF	GO Terms: Biological Process
7,76	Dal80p	• transcription • regulation of nitrogen utilization
6,43	Gln3p	• nitrogen compound metabolism • regulation of nitrogen utilization • positive regulation of transcription from Pol II promoter
6,19	Rtg3p	• transcription initiation from Pol II promoter
4,72	Pdr3p	• regulation of transcription from Pol II promoter • response to drug
4,70	Gat1p	• transcription initiation from Pol II promoter • regulation of nitrogen utilization
4,69	Rtg1p	• transcription initiation from Pol II promoter • protein localization
4,11	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
4,08	Gzf3p	• negative regulation of transcription from Pol II promoter • nitrogen compound metabolism • regulation of nitrogen utilization
4,01	Dal81p	• regulation of transcription from Pol II promoter • nitrogen utilization
3,87	Hap5p	• regulation of carbohydrate metabolism • transcription
3,84	Haa1p	• transcription initiation from Pol II promoter
3,80	Hap2p	• regulation of carbohydrate metabolism • transcription
3,79	Hap4p	• regulation of carbohydrate metabolism • transcription
3,79	Hap3p	• regulation of carbohydrate metabolism • transcription

Table 4.12. Key TFs between carbon and phosphorous limitations-anaerobic

Z-score	TF	GO Terms: Biological Process
5,37	Hap2p	• regulation of carbohydrate metabolism • transcription
5,30	Hap5p	• regulation of carbohydrate metabolism • transcription
5,28	Hap4p	• regulation of carbohydrate metabolism • transcription
5,18	Hap3p	• regulation of carbohydrate metabolism • transcription
4,53	Stp2p	• positive regulation of transcription from Pol II promoter
4,46	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
4,37	Haa1p	• transcription initiation from Pol II promoter
4,23	Yap1p	• transcription • response to oxidative stress • response to drug
4,09	Pdr3p	• regulation of transcription from Pol II promoter • response to drug
3,83	Msn2p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
3,77	Tec1p	• pseudohyphal growth • positive regulation of transcription from Pol II promoter
3,72	Cin5p	• regulation of transcription from Pol II promoter • response to salt stress • response to drug
3,64	Rtg3p	• transcription initiation from Pol II promoter

Table 4.13. Key TFs between carbon and sulfur limitations-anaerobic

Z-score	TF	GO Terms: Biological Process
5,34	Hap2p	• regulation of carbohydrate metabolism • transcription
5,32	Hap4p	• regulation of carbohydrate metabolism • transcription
5,30	Hap5p	• regulation of carbohydrate metabolism • transcription
5,26	Hap3p	• regulation of carbohydrate metabolism • transcription
4,52	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
4,06	Med8p	• Member of RNA Polymerase II transcriptional regulation mediator
3,54	Rtg1p	• transcription initiation from Pol II promoter • protein localization
3,42	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter
3,40	Rtg3p	• transcription initiation from Pol II promoter
3,18	Nrg2p	• invasive growth (sensu Saccharomyces)
3,00	Ino2p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
2,90	GFII	• binds to the upstream sequences of a number of nuclear genes coding for mitochondrial proteins and to genetic elements important for cell division in yeast
2,86	Leu3p	• regulation of transcription from Pol II promoter • leucine biosynthesis

Table 4.14. Key TFs between nitrogen and phosphorous limitations-anaerobic

Z-score	TF	GO Terms: Biological Process
8,56	Dal80p	• transcription • regulation of nitrogen utilization
7,57	Gln3p	• nitrogen compound metabolism • regulation of nitrogen utilization • positive regulation of transcription from Pol II promoter
6,22	Stp2p	• positive regulation of transcription from Pol II promoter
6,18	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
6,01	Gat1p	• transcription initiation from Pol II promoter • regulation of nitrogen utilization
5,92	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
5,16	Dal81p	• regulation of transcription from Pol II promoter • nitrogen utilization
4,71	Msn4p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
4,42	Skn7p	• transcription • response to osmotic stress • response to oxidative stress
4,37	Met32p	• sulfur amino acid metabolism

Table 4.15. Key TFs between nitrogen and sulfur limitations-anaerobic

Z-score	TF	GO Terms: Biological Process
7,72	Met28p	• sulfur amino acid biosynthesis • regulation of transcription from Pol II promoter • regulation of sulfur metabolism
7,71	Dal80p	• transcription • regulation of nitrogen utilization
7,70	Met32p	• sulfur amino acid metabolism
7,69	Met4p	• sulfur amino acid metabolism • positive regulation of transcription from Pol II promoter
7,13	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
6,75	Met31p	• sulfur amino acid metabolism • regulation of transcription
6,18	Gln3p	• nitrogen compound metabolism • regulation of nitrogen utilization • positive regulation of transcription from Pol II promoter
5,80	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
5,53	Ino2p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
5,40	Dal81p	• regulation of transcription from Pol II promoter • nitrogen utilization

Table 4.16. Key TFs between phosphorous and sulfur limitations-anaerobic

Z-score	TF	GO Terms: Biological Process
8,10	Met28p	• sulfur amino acid biosynthesis • regulation of transcription from Pol II promoter • regulation of sulfur metabolism
8,05	Met4p	• sulfur amino acid metabolism • positive regulation of transcription from Pol II promoter
7,94	Met32p	• sulfur amino acid metabolism
6,86	Met31p	• sulfur amino acid metabolism • regulation of transcription
6,09	Gcn4p	• regulation of transcription from Pol II promoter • amino acid biosynthesis
5,76	Stp2p	• positive regulation of transcription from Pol II promoter
5,42	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
5,30	Pho4p	• phosphate metabolism • cellular response to phosphate starvation
4,35	Yap1p	• transcription • response to oxidative stress • response to drug
4,31	Msn2p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress

4.2.4. Key TFs Responsive to Deletion of the Genes *MIG1* and *MIG2*

The algorithm was further implemented for analysis of transcription data from a Δ*mig1* mutant, Δ*mig1mig2* double mutant and a wild-type strain (Klein *et al*., 1999). Mig1p and Mig2p are both glucose repressors of many genes, and despite their functional redundancy, several significant differences between Mig1p and Mig2p have been reported in the literature (Lutfiyya *et al*., 1998). In the first part of this study, a *MIG1* gene deleted strain and a wild-type strain were compared in order to determine the TFs responsive to the deletion of the MIG1 gene and consequently responsive to the abscence of the TF Mig1p. Secondly, both *MIG1* and *MIG2* genes deleted strain and the wild-type strain were compared so that TFs responsive to the deletion of both of these genes could be identified. In addition, by comparing the two sets of results obtained, it was possible to determine the TFs which were responsive to the deletion of *MIG2* but not of *MIG1*. When the Δ*mig1*

mutant strain and the wild-type strain were compared, Mig2p was identified as the top key TF (Table 4.17). This is a meaningful result since it has been suggested that Mig2p may act in a redundant fashion with Mig1p (SGD, 2005). Therefore, it makes sense that Mig2p becomes very active and takes charge in the abscence of Mig1p. However Hap2/3/4/5p complex was not placed among the top 10 highest scoring TFs and was therefore not identified as key TF in this data set (Table 4.17). This result was expected because it has been reported that a *mig1* mutation did not affect the regulation of *HAP4* even though a Mig1p binding site was present in the *HAP4* promoter (Gancedo, 1998). When Δ*mig1mig2* double mutant strain and the wild-type strain were analyzed, this time Hap2/3/4/5p complex emerged as the top key TF (Table 4.18). This result, along with the previous one, suggests that *MIG2* but not *MIG1* has an effect on Hap2/3/4/5p activity. Indeed, it has been suggested that perhaps Mig2p could be involved in the mechanism which regulates *HAP4* (Gancedo, 1998). Our results strengthen this hypothesis. Additionally, it has been reported that Δ*mig1mig2* double mutant exhibited a significantly higher respiratory capacity than the wild type (Klein *et al*., 1999). The cause behind this phenomenon may be, as found here, the increased activity of Hap2/3/4/5p, global regulator of respiratory gene expression.

Topological measures of *MIG1* and *MIG1MIG2* deleted networks indicated that removal of these genes from the transcriptional regulatory network did not affect the network topology significantly (Table 4.20). The reason for this result is that both *MIG1* and *MIG2* make relatively small number of connections in the network; hence, they are not hubs in the network and do not affect the network topology significantly. In fact, their removal resulted in small decreases in the *in-degree* ($<k_{in}>$) and the *out-degree* ($<k_{out}>$) values (Table 4.20). The exponential and power-law behaviors were maintained in the deletion networks for the arriving and the departing connectivity distributions, respectively (Figures 4.8 – 4.11). The exponent γ slightly decreased in the *MIG1MIG2* deleted network while the exponent β remained the same in the deletion networks as in the original network (Table 4.20).

The *path length (l)* increased slightly and gradually as *MIG1* and *MIG1MIG2* genes were removed from the network whereas the *diameter (d)* of the original network did not change upon removal of *MIG1* and *MIG1MIG2* genes from the network (Table 4.20).

Table 4.17. Key TFs responsive to the deletion of *MIG1* gene

Z-score	TF	GO Terms: Biological Process
3,11	Mig2p	• glucose metabolism • regulation of transcription from Pol II promoter
1,97	Arg81p	• arginine metabolism • negative regulation of calcium ion-dependent exocytosis
1,92	Mal13p	• carbohydrate metabolism • regulation of transcription, DNA-dependent
1,80	Stp1p	• tRNA splicing • positive regulation of transcription from Pol II promoter
1,71	Uga3p	• regulation of transcription from Pol II promoter • nitrogen utilization
1,66	BUF	• activator or repressor involved in replication
1,52	Gal4p	• galactose metabolism • regulation of transcription, DNA-dependent
1,48	Gzf3p	• negative regulation of transcription from Pol II promoter • nitrogen compound metabolism • regulation of nitrogen utilization
1,48	Gal80p	• Transcriptional regulator involved in the repression of GAL genes in the absence of galactose
1,44	Smp1p	• Positive regulation of transcription from Pol II promoter

Table 4.18. Key TFs responsive to the deletion of both *MIG1* and *MIG2* genes

Z-score	TF	GO Terms: Biological Process
7,91	Hap4p	• regulation of carbohydrate metabolism • transcription
7,80	Hap5p	• regulation of carbohydrate metabolism • transcription
7,72	Hap2p	• regulation of carbohydrate metabolism • transcription
7,66	Hap3p	• regulation of carbohydrate metabolism • transcription
5,35	Med8p	• Member of RNA Polymerase II transcriptional regulation mediator
5,02	Cat8p	• gluconeogenesis • positive regulation of transcription from Pol II promoter
4,46	Ino2p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
4,01	Ino4p	• phospholipid biosynthesis • positive regulation of transcription from Pol II promoter
3,90	Phd1p	• pseudohyphal growth
3,89	Ppr1p	• regulation of transcription, DNA-dependent • uracil biosynthesis
3,78	Nrg2p	• invasive growth (sensu Saccharomyces)
3,21	Msn4p	• replicative cell aging • age-dependent response to oxidative stress during chronological cell aging • regulation of transcription, DNA-dependent • response to stress
3,20	Oaf1p	• fatty acid metabolism • peroxisome organization and biogenesis • positive regulation of transcription

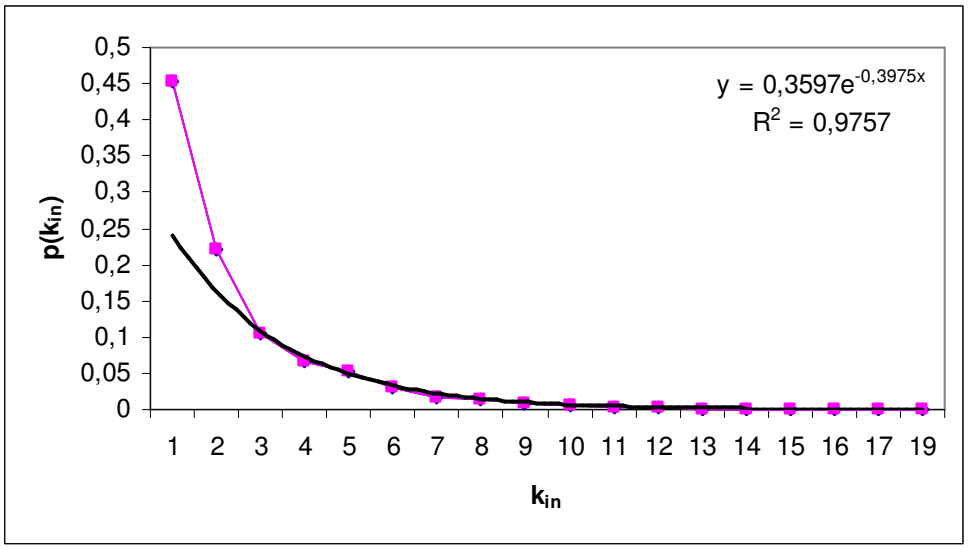

Figure 4.8. Exponential behavior of Δ*MIG1* network arriving connectivity

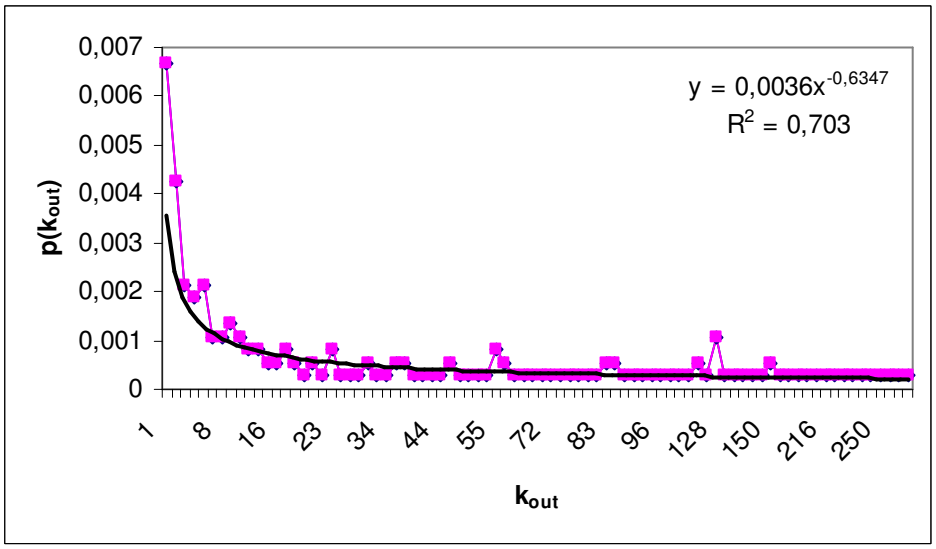

Figure 4.9. Power-law behavior of Δ*MIG1* network departing connectivity

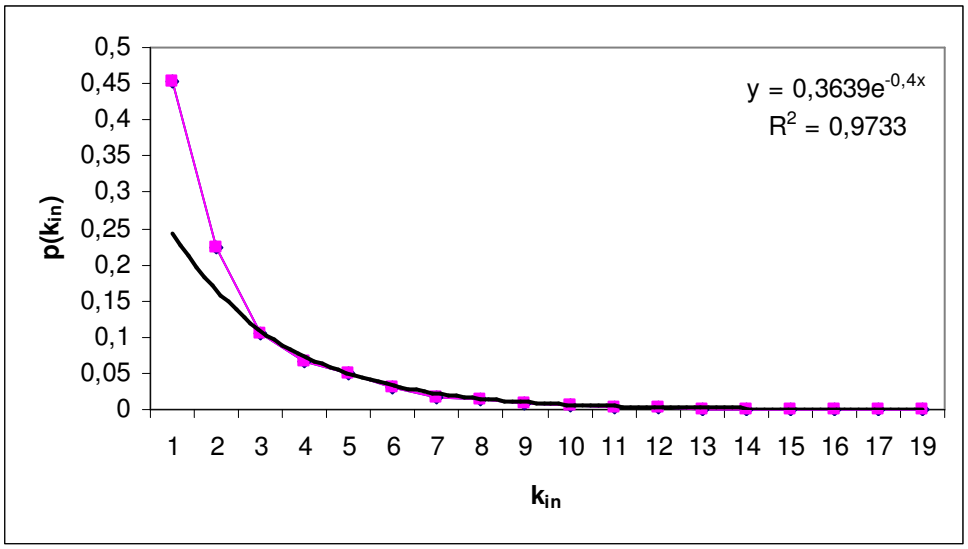

Figure 4.10. Exponential behavior of Δ*MIG1MIG2* network arriving connectivity

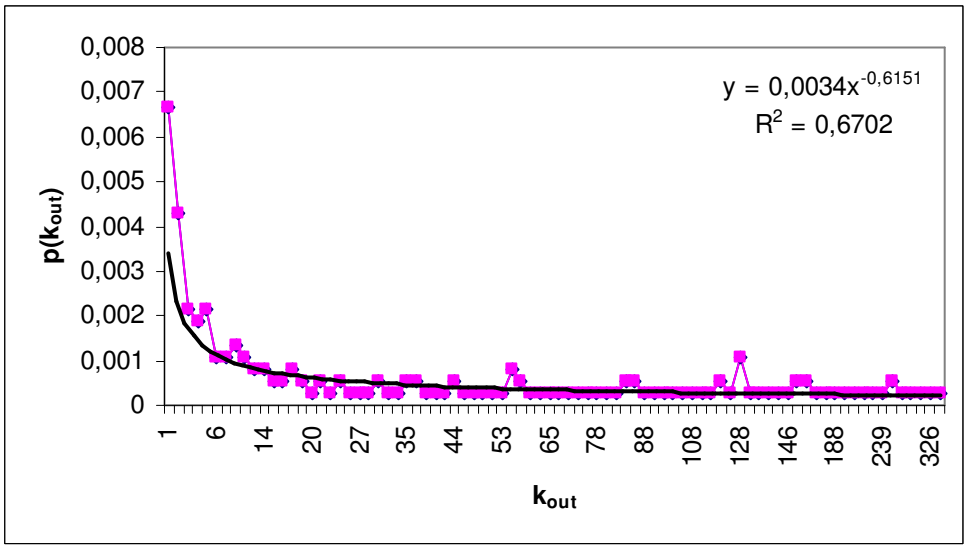

Figure 4.11. Power-law behavior of Δ*MIG1MIG2* network departing connectivity

4.3. Identification of Perturbation-Responsive Subnetworks

Perturbation-responsive subnetwoks were identified by interconnecting the key TFs and their differentially expressed target genes responsive to the same perturbations. The genes whose fold changes (FC) (*i.e.* the ratio of their expression levels) were greater than $|2|$ were accepted as differentially expressed. The algorithm here reported is used for the identification of key TFs, and the differentially expressed genes regulated by key TFs were selected from gene expression data provided by Tai *et al*., 2005.

The oxygen responsive subnetwork under carbon limitation regime was constructed as an illustration (Figure 4.12). The subnetwork consists of 10 TFs (key TFs) and 87 target genes (Table 4.19) activated when passing from anaerobic to aerobic conditions under carbon limitation. As expected, the largest portion of the genes (16 per cent) in the subnetwork belongs to the MIPS (Munich Information Center for Protein Sequences) functional category 'respiration' (Figure 4.13). The second largest categories are 'transported compounds' (14 per cent) and 'unknown proteins' (14 per cent) (Figure 4.3). It is known that in oxygen-deprived environments, yeast uses compounds such as sulfate, carbonate or nitrate in place of oxygen (Yesmag, 2004). Therefore, it is quite logical that when passing from anaerobic to aerobic conditions the expression levels of the genes related to transported compounds change. Of great interest are genes that have unknown functions: a large portion of the genes in the subnetwork were classified as 'unknown proteins' by MIPS (Figure 4.13) and they are displayed in bold in Table 4.19. The classification of these genes in specified subnetworks together with TFs with known functions can help the prediction of their functions. The fourth largest portion of the genes belongs to the category 'electron transport and membrane-associated energy', which is clearly related to the respiration (Figure 4.13). Stress response (9 per cent), lipid metabolism (9 per cent) and carbohydrate metabolism (8 per cent) are among the top 10 important categories that the genes in the oxygen responsive subnetwork belong to (Figure 4.3). This result is totally consistent with a recent study which revealed that a large fraction of the anaerobically induced genes were involved in cell stress (approximately 1/3), carbohydrate metabolism (approximately 1/10), and lipid metabolism (approximately 1/12) (Kwast *et al*., 2002).

The topological measures of the subnetwork revealed that the *in-degree exponential exponent (β)* doubled and the *out-degree power-law exponent (γ)* halved between the original network and the subnetwork. The doubled exponent is indicative of a simpler TF combination usage in the subnetwork and the halved exponent means that the subnetwork contains fewer TF hubs (Luscombe *et al*, 2004) than the original network. The exponential and the power-law behaviors were maintained in the subnetwork for the distributions of the arriving and the departing connectivities, respectively (Figures 4.14 and 4.15). Notably, the *in-degree* and *out-degree* values halved between the original network and the subnetwork,

since the number of TFs and genes in the subnetwork is smaller compared to the original network (Table 4.20).

Table 4.19. Genes included in the oxygen-responsive subnetwok

Data set	Genes in the subnetwork. Bold names represent unknown genes.
ANA-C-AER-C	*AAC3, ADH5, AGX1, ANT1, ARE1, ATF2, ATP20, ATP7, BNA2, CBP4, COR1, COX12, COX4, COX5A, COX6, COX7, CRC1, CYB2, CYC1, CYT1, DAN1,* ***DAN2****, DAN3,* ***DAN4****, EMI2, FET4, FMP13, GRE2, GSY1, GUT2, HEM13, HES1, IZH2,* ***JID1****, KGD2, LSP1, MBA1, MCH4, MCR1, MDG1, NDI1, OSH7, PAU1, PAU4, PAU5, PAU6, PDR11,* ***PHM7****, PUT4, QCR2, RIM4, RIP1, ROX1, SDH1, SOD2, SOL4, SRO77, SUT1, TIR1, TIR2, TIR3, TIR4, UPC2,* ***YAL068C, YBR230C, YCR061W, YDL241W, YDR542W, YGL196W, YGL261C,*** *YGR131W,* ***YGR294W****, YHL046C,* ***YIL176C****, YJL218W,* ***YLL064C****,* ***YLR108C****, YLR168C,* ***YLR413W****,* ***YML083C****, YML087C,* ***YMR325W****, YNL274C, YOL155C, YPC1,* ***YPL272C****, YTP1*

Table 4.20. Topological measures of the constructed networks

topological measure	**networks and cellular conditions**			
	original	*Δmig1*	*Δmig1mig2*	*Subnetwork (ANA-C-AER-C)*
in-degree (<k_{in}>)	2.43	2.42	2.42	1.43
in-degree exponential exponent (β)	0.40	0.40	0.40	0.94
out-degree (<k_{out}>)	2.43	2.42	2.42	1.41
out-degree power-law exponent (γ)	0.64	0.64	0.62	0.27
path length (l)	4.25	4.27	4.28	1.00
diameter (d)	13	13	13	1

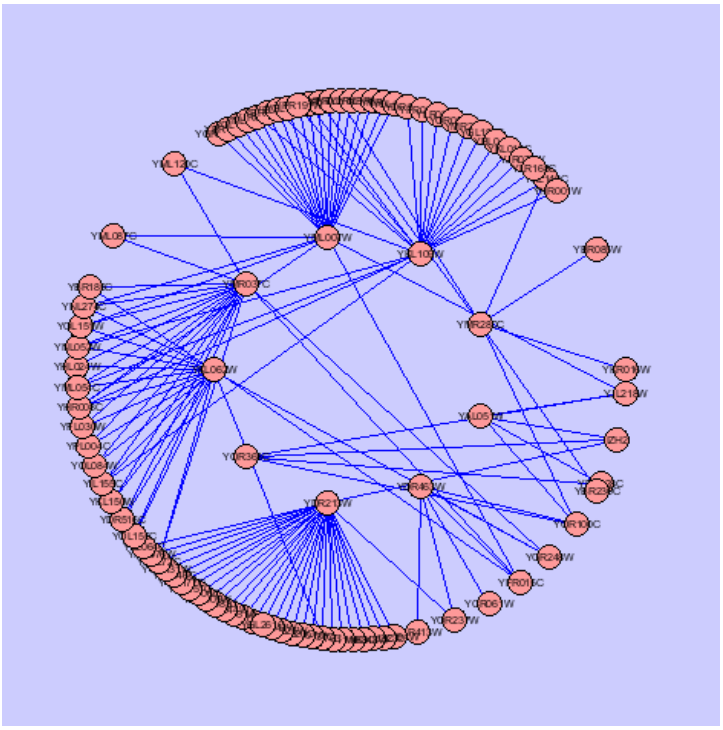

Figure 4.12. Oxygen-responsive subnetwork under carbon limitation regime

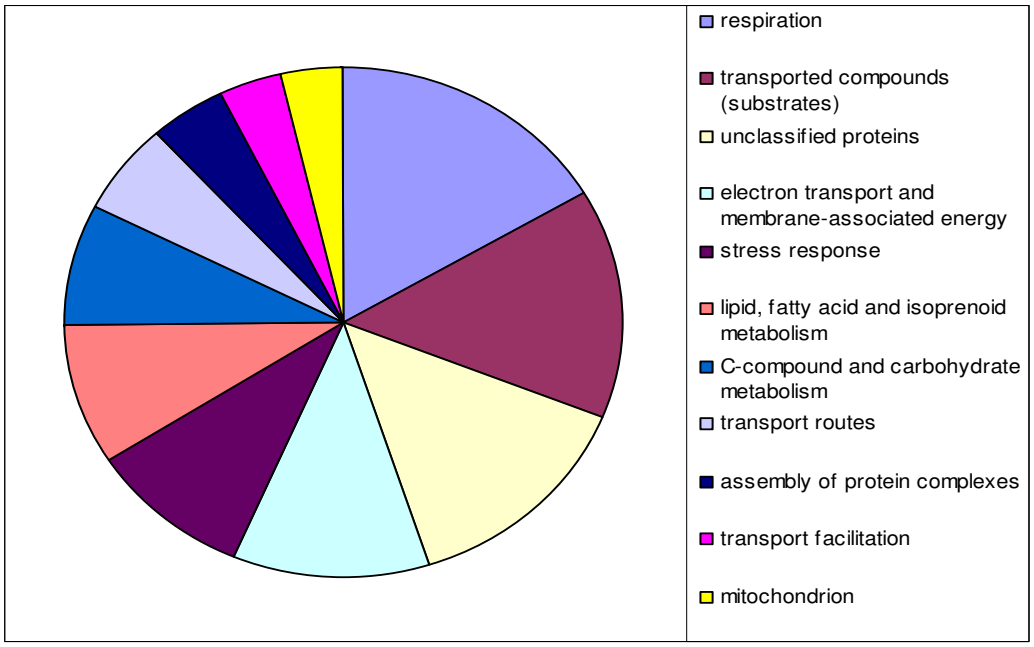

Figure 4.13. Distribution of the genes from the subnetwork into different MIPS functional categories. Top 10 categories are shown for simplicity

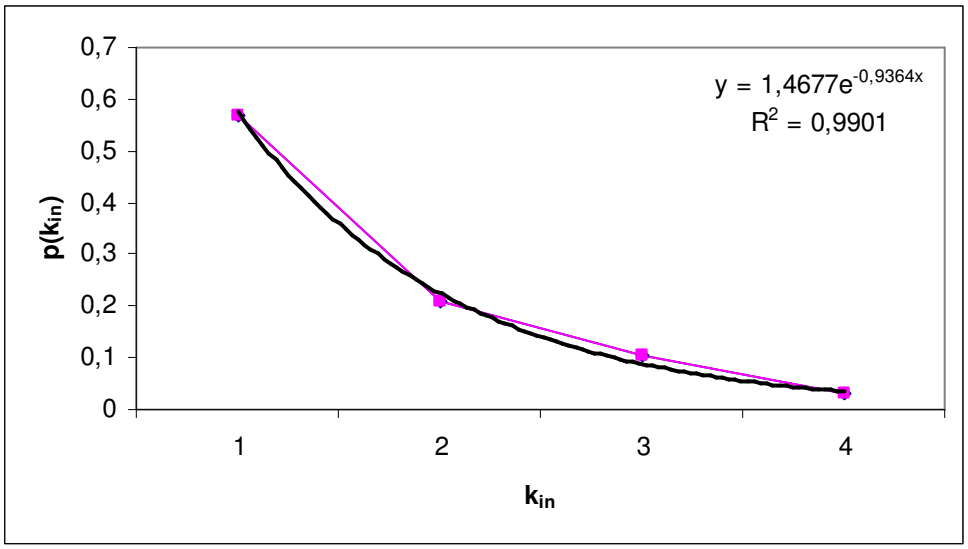

Figure 4.14. Exponential behavior of the subnetwork arriving connectivity

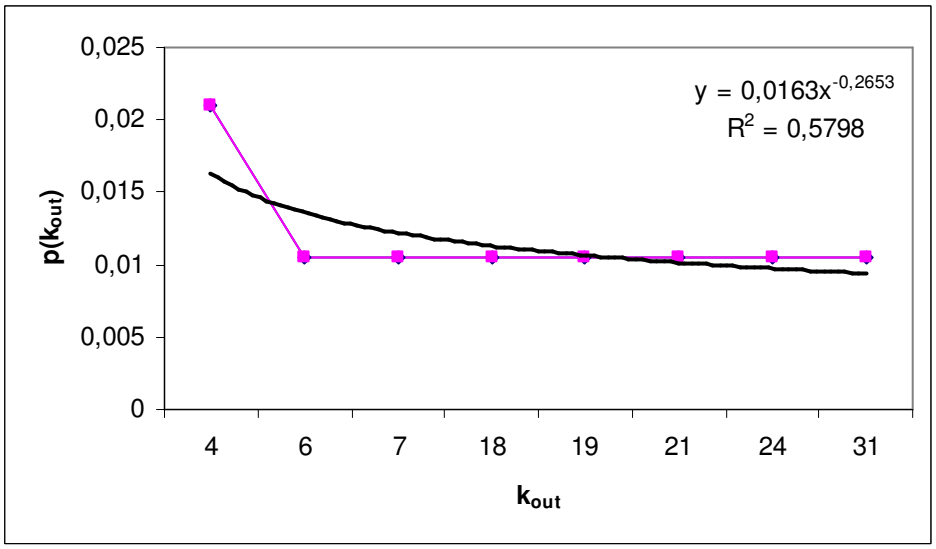

Figure 4.15. Power-law behavior of the subnetwork departing connectivity

4.4. Best Candidate Target Genes for the Key TFs

There is a very large number of genes whose promotor regions contain binding sites for TFs. However, the regulatory interactions between these genes and these TFs are not experimentally justified. These genes are the so-called “potential target genes” for these TFs, because they have potential to be truly regulated by these TFs. The reason for this potentiality is that they posses binding sites of these TFs in their promoter regions.

In the first step of this study, the potential target genes of key TFs were found in the database YEASTRACT. Next, those that responded very significantly (*i.e.* that are upregulated or downregulated very significantly) to the same perturbation used for key TF identification were selected from gene expression data according to their fold change (FC>|10|) (Tai *et al*., 2005). Here, a high threshold value for the fold change was chosen in order to restrict the number of differentially expressed genes. For example, there are 3513 genes whose promoter regions contain binding sites for the transcription factor Msn2p (Table 4.21). These 3513 genes are potential target genes for Msn2p. Since Msn2p was identified as a key TF between anaerobic and aerobic conditions under the carbon limited regime, it is expected to regulate genes which are equally differentially expressed between the same conditions. Therefore, from gene expression data, genes which are upregulated or downregulated significantly (FC>|10|) were selected. The 3513 potential target genes listed in the database YEASTRACT was then searched for these genes. Finally, the ones which are both among the potential target genes and the differentially expressed genes were identified as best candidates truly being regulated by Msn2p.

The best candidate target genes for each key TF are listed in Tables 4.21 - 4.36. These results should provide useful start points for further experimental investigations of regulatory interactions of key TFs or the selected best candidate target genes.

Table 4.21. The best candidate target genes for the set ANA-C-AER-C

Key TF	**# of Potential Target Genes**	**Best Candidate Target Genes of the Key TFs**
Msn2p	3513	*AAC3, ADI1, AGX1, AQY2, AUS1, CRC1, CYB2, DAN1, ECM13, FAA1, FET4, FMP23, GUT2, HEM13, HMX1, LSB6, NDE1, PNS1, PUT4, RIM4, SOL4, SUE1, TIR3, TIR4, YBR230c, YLL053c, YLR312c, YLR413w*
Msn4p	3513	*AAC3, ADI1, AGX1, AQY2, AUS1, CRC1, CYB2, DAN1, ECM13, FAA1, FET4, FMP23, GUT2, HEM13, HMX1, LSB6, NDE1, PNS1, PUT4, RIM4, SOL4, SUE1, TIR3, TIR4, YBR230c, YLL053c, YLR312c, YLR413w*
Stp1p	898	*ADI1, AGX1, AQY2, FAA1, HEM13, LSB6, RIM4, SUE1, TIR4, YLL053c, YSR3*
Yap1p	1291	*AGX1, AQY2, AUS1, ECM13, FAA1, FMP23, LSB6, MUC1, TIR3, YHL042w*
Upc2p	793	*ADI1, DAN1, HMX1, TIR1, TIR3, TIR4, YLR312c, YLR413w, YSR3*
Hap2/3/4/5p	1140	*CYB2, ECM13, GRE2, GUT2, HEM13, NDE1, PNS1*
Pip2p	259	*ECM13, PUT4, RIM4, TIR1*
Oaf1p	259	*ECM13, PUT4, RIM4, TIR1*
Cat8p	320	*LSB6, SUE1*

Table 4.22. The best candidate target genes for the set ANA-N-AER-N

Key TF	**# of Potential Target Genes**	**Best Candidate Target Genes of the Key TFs**
Upc2p	793	*ATF2, DAN1, HES1, HMX1, TIR1, TIR3, TIR4, YLR413w, YOL161c, YSR3*
Stp1p	898	*HEM13, SUE1, TIR4, YOL161c, YSR3*
Stp2p	898	*HEM13, SUE1, TIR4, YOL161c, YSR3*
Hap2/3/4/5p	1140	*CYB2, ECM13, HEM13, IZH4, NDE1*
Yap1p	1291	*AUS1, ECM13, MUC1, TIR3*
Rox1p	28	*ANB1, CYC1, HEM13*
Oaf1p	259	*ECM13, PUT4, TIR1*
Pip2p	259	*ECM13, PUT4, TIR1*

Table 4.23. The best candidate target genes for the set ANA-P-AER-P

Key TF	**# of Potential Target Genes**	**Best Candidate Target Genes of the Key TFs**
Mot3p	30914	*AAC3, AGX1, ANB1, AQY2, AUS1, COX5a, CYB2, CYC1, CYT1, DAN1, DAN2, DAN3, DAN4, ECM13, FAA1, HEM13, HES1, HMX1, IZH4, LSB6, MUC1, NDE1, PAU1, PAU3, PAU4, PAU6, PLB2, PUT4, SML1, SOD2, SUE1, TIR1, TIR3, TIR4, YDR542w, YGL261c, YGR131w, YGR294w, YHL046c, YIL176c, YLL053c, YLL064c, YLR413w, YMR325w, YOL155c, YOL161c, YSR3*
Rgt1p	4285	*AAC3, AGX1, ANB1, AQY2, COX5a, CYB2, CYC1, DAN1, DAN2, DAN3, DAN4, ECM13, HEM13, HMX1, IZH4, MUC1, NDE1, PAU3, PAU4, PAU6, PLB2, PUT4, SML1, SOD2, SUE1, TIR1, TIR3, TIR4, YGL261c, YGR294w, YHL046c, YLL053c, YLR413w, YMR325w, YOL155c, YSR3*
Upc2p	793	*DAN1, DAN2, DAN3, DAN4, HES1, HMX1, PAU1, PAU3, PAU4, PAU6, PLB2, TIR1, TIR3, TIR4, YDR542w, YGL261c, YGR294w, YIL176c, YLL064c, YLR413w, YMR325w, YOL155c, YOL161c, YSR3*
Yap1p	1291	*AGX1, AQY2, AUS1, DAN3, ECM13, FAA1, LSB6, MUC1, SOD2, TIR3, YGR294w, YLL064c, YOL155c*
Hap2p	1140	*CYB2, ECM13, HEM13, IZH4, NDE1, PLB2, SOD2, YMR325w*
Rox1p	28	*ANB1, CYC1, HEM13*
Met32p	439	*NDE1, TIR3*
Cat8p	320	*LSB6, SUE1*

Table 4.24. The best candidate target genes for the set ANA-S-AER-S

Key TF	**# of Potential Target Genes**	**Best Candidate Target Genes of the Key TFs**
Mot3p	30914	*AAC3, ANB1, AQY2, ATF2, AUS1, COX5a, CYB2, CYC1, CYT1, DAN1, DAN3, ECM13, FAA1, FET4, HEM13, HES1, HMX1, IZH4, MUC1, NDE1, PAU1, PAU3, PAU4, PAU6, SML1, SUE1, TIR1, TIR3, TIR4, YDR542w, YGR131w, YGR294w, YHL042w, YHL046c, YIL176c, YLL053c, YLL064c, YLR413w, YMR325w, YOL161c*
Rtg3p	8344	*AQY2, ATF2, AUS1, COX5a, CYB2, CYT1, DAN1, DAN3, ECM13, FAA1, HES1, HMX1, IZH4, MUC1, NDE1, PAU1, PAU3, PAU4, SUE1, TIR1, TIR3, TIR4, YDR542w, YGR131w, YGR294w, YHL042w, YHL046c, YIL176c, YLL053c, YLL064c, YMR325w*
Upc2p	793	*ATF2, DAN1, DAN3, HES1, HMX1, PAU1, PAU3, PAU4, PAU6, TIR1, TIR3, TIR4, YDR542w, YGR294w, YIL176c, YLL064c, YLR413w, YMR325w, YOL161c*
Yap1p	1291	*AQY2, AUS1, DAN3, ECM13, FAA1, MUC1, TIR3, YGR294w, YHL042w, YLL064c*
Hap1p	398	*ANB1, CYC1, CYT1, FAA1, HEM13, HMX1, IZH4, TIR4*
Cad1p	1291	*AQY2, AUS1, DAN3, FAA1, MUC1, TIR3, YGR294w, YLL064c*
Pdr3p	439	*HEM13, PAU3, PAU4, SUE1, TIR4, YGR131w, YLR413w*
Hap2p	1140	*CYB2, ECM13, HEM13, IZH4, NDE1, YMR325w*
Rox1p	28	*ANB1, CYC1, HEM13*

Table 4.25. The best candidate target genes for the set AER-C-N

Key TF	**# of Potential Target Genes**	**Best Candidate Target Genes of the Key TFs**
Gln3p	4465	*JEN1, PDH1, WSC4, YER188w, YIL057c, YMR206w*
Stp2p	898	*FMP43, HXT3, JEN1, YIL057c, YMR206w*
Mcm1p	6250	*FMP43, PDH1, WSC4, YER188w*
Gat1p	2592	*WSC4, YER188w, YIL057c, YMR206w*
Pip2p	259	*HXT3, YIL057c, YMR206w*
Oaf1p	259	*HXT3, YIL057c, YMR206w*
Cat8p	320	*JEN1*

Table 4.26. The best candidate target genes for the set AER-C-P

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Msn2p	3513	*HXT1, HXT3, ISF1, JEN1, MAL32, MIG2, PDH1, RBF9, SUC2, TPO2, WSC4, YER188w, YMR206w*
Msn4p	3513	*HXT1, HXT3, ISF1, JEN1, MAL32, MIG2, PDH1, RBF9, SUC2, TPO2, WSC4, YER188w, YMR206w*
Rtg1p	8344	*HXT1, HXT3, ISF1, JEN1, MAL32, MIG2, RBF9, SUC2, TPO2, WSC4, YER188w, YIL057c*
Stp2p	898	*HXT3, ISF1, JEN1, YIL057c, YMR206w*
Stp1p	898	*HXT3, ISF1, JEN1, YIL057c, YMR206w*
Pip2p	259	*HXT3, MIG2, YIL057c, YMR206w*
Oaf1p	259	*HXT3, MIG2, YIL057c, YMR206w*
Pdr3p	439	*HXT3, MIG2, YIL057c*
Mig2p	23	*MAL32, SUC2*

Table 4.27. The best candidate target genes for the set AER-C-S

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Msn4p	3513	*HXK1, HXT1, HXT3, ISF1, JEN1, MAL32, SOL1, YMR206w*
Gcn4p	5970	*HXT1, JEN1, MAL32, YIL057c, YMR206w*
Pdr3p	439	*HXT3, YIL057c*
Cat8p	320	*ISF1, JEN1*
Yap1p	1291	*HXT1*
Met32p	439	*YMR206w*
Mig2p	23	*MAL32*

Table 4.28. The best candidate target genes for the set AER-N-P

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Rtg3p	8344	*CPS1, DAL2, DAL4, DAL5, DAL80, MLS1, OPT2, PUT1, VBA1*
Gln3p	4465	*CPS1, DAL2, DAL4, DAL5, DAL80, OPT2, PUT1, VBA1*
Gat1p	2592	*CPS1, DAL2, DAL4, DAL5, DAL80, OPT2, PUT1, VBA1*
Gcn4p	5970	*CPS1, DAL2, DAL5, DAL80*
Pho4p	2345	*DAL80, OPT2, PUT1*
Stp1p	898	*DAL80, MLS1*
Stp2p	898	*DAL80, MLS1*
Ste12p	1478	*OPT2*
Hap1p	398	*PUT1*

Table 4.29. The best candidate target genes for the set AER-N-S

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Gln3p	4465	*CPS1, DAL1, DAL2, DAL4, DAL5, DAL80, DUR3, GAP1, OPT2, PUT1, VBA1*
Gcn4p	5970	*CPS1, DAL2, DAL5, DAL80, DUR3, GAP1, YLR053c*
Yap1p	1291	*CPS1, DAL2, DUR3, GNP1, YLR053c*
Dal81p	100	*DAL1, DAL4, MLS1*
Stp1p	898	*DAL80, MLS1*
Pdr3p	439	*DAL2*
Met32p	439	*MLS1*
Ino4p	272	*DAL2*
Ino2p	240	*DAL2*

Table 4.30. The best candidate target genes for the set AER-P-S

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Gcn4p	5970	*ICY1, PHM6, PHO3, SPL2, SUL1*
Msn4p	3513	*GIT1, ICY1, PHM8, PHO84, SPL2*
Msn2p	3513	*GIT1, ICY1, PHM8, PHO84, SPL2*
Met4p	470	*ICY1, PHO84, SPL2*
Yap1p	1291	*PHO3, SUL1*
Stp2p	898	*PHO11, SPL2*
Pdr3p	439	*ICY1, SPL2*
Stp1p	898	*PHO11, SPL2*
Met32p	439	*GFD2*

Table 4.31. The best candidate target genes for the set ANA-C-N

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Rtg1p	8344	*HXT1, HXT3, HXT4, MAL32, MRK1, SUC2, YIL057c*
Rtg3p	8344	*HXT1, HXT3, HXT4, MAL32, MRK1, SUC2, YIL057c*
Pdr3p	439	*HXT3, YIL057c*
Stp1p	898	*HXT3, YIL057c*
Gln3p	4465	*HXT4, YIL057c*
Gzf3p	2592	*HXT4, YIL057c*
Gat1p	2592	*HXT4, YIL057c*
Hap2p	1140	*HXT1*

Table 4.32. The best candidate target genes for the set ANA-C-P

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Rtg3p	8344	*HXT3, HXT4, MAL11, MAL32, MRK1, SUC2, TPO2, YIL057c*
Tec1p	5584	*HXT3, HXT4, MAL11, MAL32, MRK1, SUC2, TPO2, YIL057c*
Msn2p	3513	*HXT3, HXT4, MAL32, MRK1, SUC2, TPO2*
Stp2p	898	*HXT3, YIL057c*
Pdr3p	439	*HXT3, YIL057c*
Stp1p	898	*HXT3, YIL057c*
Yap1p	1291	*MRK1*

Table 4.33. The best candidate target genes for the set ANA-C-S

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Rtg3p	8344	*HXT1, HXT3, HXT4, ISF1, MAL11, MAL32, MRK1, SUC2, TPO2, YIL057c*
Rtg1p	8344	*HXT1, HXT3, HXT4, ISF1, MAL11, MAL32, MRK1, SUC2, TPO2, YIL057c*
Gcn4p	5970	*HXT1, HXT4, MAL11, MAL32, MRK1, TPO2, YIL057c, YMR206w*
Hap2p	1140	*HXT1, ISF1, YMR206w*
Mig2p	23	*MAL32, SUC2*

Table 4.34. The best candidate target genes for the set ANA-N-P

Key TF	# of Potential Target Genes	Best Candidate Target Genes of the Key TFs
Gat1p	2592	*DAL2, DAL5, DUR1,2, GAP1, MEP2, PUT1, VBA1*
Gln3p	4465	*DAL2, DAL5, DUR1,2, GAP1, MEP2, PUT1, VBA1*
Gcn4p	5970	*DAL2, DAL5, DUR1,2, GAP1, MEP2*
Msn4p	3513	*DUR1,2, GAP1, MLS1*
Dal81p	100	*DUR1,2, MLS1*
Stp2p	898	*MLS1*
Stp1p	898	*MLS1*
Skn7p	2366	*PUT1*
Met32p	439	*MLS1*

Table 4.35. The best candidate target genes for the set ANA-N-S

Key TF	**# of Potential Target Genes**	**Best Candidate Target Genes of the Key TFs**
Gln3p	4465	*DAL5, DAL80, GAP1, MEP2, OPT2, PUT1, VBA1*
Gcn4p	5970	*DAL5, DAL80, GAP1, MEP2*
Met4p	470	*GNP1, OPT2, PUT1*
Stp1p	898	*DAL80, MLS1*
Met32p	439	*MLS1*
Met31p	439	*MLS1*
Dal81p	100	*MLS1*

Table 4.36. The best candidate target genes for the set ANA-P-S

Key TF	**# of Potential Target Genes**	**Best Candidate Target Genes of the Key TFs**
Pho4p	2345	*GIT1, PHM6, PHO11, PHO84, PHO89, SPL2, VTC3*
Gcn4p	5970	*PHM6, PHO3, PHO89, SPL2, SUL1, VTC3*
Msn2p	3513	*GIT1, PHO84, SPL2*
Yap1p	1291	*PHO3, SUL1*
Stp2p	898	*PHO11, SPL2*
Met4p	470	*PHO84, SPL2*
Stp1p	898	*PHO11, SPL2*

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

It is concluded that perturbation-responsive key transcription factors (transcription factors around which most significant transcriptional changes occur when a perturbation is introduced) can be identified by the hypothesis-driven approach used in this study. The approach was implemented in this study for identification of transcription factors responsive to genetic and environmental perturbations.

Mig2p was identified as the top key transcription factor responding to the deletion of the gene *MIG1*. Hap2/3/4/5p complex was found to be the transcription factor which responded the most significantly to the deletion of both *MIG1* and *MIG2* genes. These results suggest that *MIG2*, but not *MIG1*, is involved in the mechanism which regulates HAP genes. As for environmental perturbations, oxygen-responsive transcription factors under four macronutrient limitation regimes and macronutrient-responsive transcription factors under aerobic and anaerobic conditions were identified. Hap2/3/4/5p complex, Upc2p and Yap1p were identified as trancription factors responding significantly to oxygen availability irrespective of nutrient limitation regime. Meaningful results were also obtained for nutrient limitation-responsive transcription factors. For example, Gln3p and Gln80p, two TFs involved in nitrogen utilization process, were identified as key TFs only when passing from nitrogen limitation regime to any other nutrient limitation regime.

It was furthermore showed that once the key transcription factors are identified by this method, perturbation-responsive subnetworks can be constructed by interconnecting the key transcription factors and their target genes responding significantly to the same perturbation. For illustration, oxygen-responsive subnetwork was constructed. Very large portion of the genes included in the subnetwork were classified as ‘unknown proteins’ by MIPS. We anticipate that the method will be useful in predicting functions for unknown or poorly characterized genes included in a specific subnetwork after a known perturbation is introduced to the system. Likewise, the effects of unknown perturbations, such as deletion

of a gene with unknown function or drug exposure, can be predicted from key transcription factors and subnetworks activated upon perturbation.

The best candidate target genes for each key transcription factor were also identified by selecting among the genes, whose promoter sequences match with the binding site of the key TF, the ones which are significantly upregulated or downregulated after specific perturbations. We believe the genes identified as best candidate target genes for key transcription factors should provide useful start points for further experimental investigations of regulatory interactions of these genes or key transcription factors.

5.2. Recommendations

The transcritional regulatory network constructed may also be represented as a unipartite undirected graph, in which only genes are represented as the nodes and the genes regulated by a common transcription factor are connected to each other. Consequently, the unipartite graph can be integrated with multidimensional data (gene expression is measured over a time course or multiple strains are analysed). The use of multidimensional data would enable the identification of condition-specific key transcription factors during specific periods of time, such as aging in yeast, diauxic shift time course, temporal analysis of sporulation, heat shock time course etc.

When determining the differentially expressed genes in the subnetwork and the best candidate target genes analyses, *p* values may be used instead of fold change since triplicate data were used and *p* values can be readily calculated from these data.

A different statistical method, other than the *t* test, may be tried so that the method proposed in this study for the identification of the perturbation-responsive key transcription factors can be used when replicate data are not available.

The occurrence of *network motifs*, which are compact, specific patterns of inter-connection between transcription factors and targets, may be calculated in the genome-scale transcriptional regulatory network or in the subnetworks. Motif usage between different conditions or during a time course may be investigated.

REFERENCES

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts and J. D. Watson, 1994, *Molecular Biology of the Cell,* Garland Publishing, New York.

Barabasi, A. L. and Z. N. Oltvai, 2004, "Network Biology: Understanding the Cell's Functional Organization", *Nature Genetics,* Vol. 5, pp. 101-113.

Bar-Joseph, Z., G. K. Gerber, T. I. Lee, N. J. Rinaldi, J. Y. Yoo, F. Robert, D. B. Gordon, E. Fraenkel, T. S. Jaakkola, R. A. Young, D. K. Gifford, 2003, "Computational Discovery of Gene Modules and Regulatory Networks", *Nature Biotechnology,* Vol. 21, pp. 1337-1342.

Bergmann, S., J. N. Ihmels, N. Barkai, 2003, "Iterative Signature Algorithm for the Analysis of Large-Scale Gene Expression Data", *Physical Review E,* Vol. 67, 031902.

Blaiseau, P. L., A. D. Isnard, Y. Surdin-Kerjan and D. Thomas, 1997, "Met31p and Met32p, Two Related Zinc Finger Proteins, Are Involved in Transcriptional Regulation of Yeast Sulfur Amino Acid Metabolism", *Molecular and Cellular Biology*, Vol. 17, pp. 3640-3648.

Bourgarel, D., C. C. Nguyen and M. Bolotin-Fukuhara, 1999, "HAP4, the Glucose-Repressed Regulated Subunit of the HAP Transcriptional Complex Involved in the Fermentation-Respiration Shift, Has a Functional Homologue in the Respiratory Yeast *Kluyveromyces lactis*", *Molecular Microbiology*, Vol. 31, pp. 1205-1215.

Bricmont, P. A., J. R. Daugherty and T. R. Cooper, 1991, "The DAL81 Gene Product Is Required for Induced Expression of Two Differently Regulated Nitrogen Catabolic Genes in *Saccharomyces cerevisiae*", *Molecular and Cellular Biology*, Vol. 11, pp. 1161-1166.

Buchler, N.E., U. Gerland and T. Hwa, 2003, "On Schemes of Combinatorial Transcription Logic", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 100, pp. 5136-5141.

Cases, I., V. de Lorenzo and C. A. Ouzounis, 2003, "Transcription Regulation and Environmental Adaptation in Bacteria", *Trends in Microbiology*, Vol. 11, pp. 248-253.

Chan, T. F., P. G. Bertram, W. Ai and X. F. Zheng, 2001, "Regulation of APG14 Expression by the GATA-Type Transcription Factor Gln3p", *Journal of Biological Chemistry,* Vol. 276, pp. 6463-6467.

Csank, C., M. C. Costanzo, J. Hirschman, P. Hodges, J. E. Kranz, M. Mangan, K. O'Neill, L. S. Robertson, M. S. Skrzypek and J. Brooks, 2002, "Three Yeast Proteome Databases: YPD, PombePD, and CalPD (MycoPathPD)", *Methods in Enzymology*, Vol. 350, pp. 347-373.

Dhaseieer, P., S. Liang and R. Somogoyi, 2000, "Genetic Network Inference: from Co-Expression Clustering to Reverse Engineering", *Bioinformatics*, Vol. 16, pp. 707-726.

Eisenstein, M., 2005, "Microarrays: A Plan for When the ChIPs Are Down", *Nature Methods*, Vol.5.

Gancedo, J. M., 1998, "Yeast Carbon Catabolite Repression", *Microbiology and Molecular Biology Reviews : MMBR,* Vol. 62, pp. 334-361.

Gardner, T.S., D. Bernardo, D. Lorenz and J. J. Collins, 2003, "Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling", *Science,* Vol. 301, pp. 102-105.

Guelzim, N., S. Bottani, P. Bourgine and F. Kepes, 2002, "Topological and Causal Structure of the Yeast Transcriptional Regulatory Network", *Nature Genetics,* Vol. 31, pp. 60-63.

Gutierrez-Rios, R. M., D. A. Rosenblueth, J. A. Loza, A. M. Huerta, J. D. Glasner, F. R. Blattner and J. Collado-Vides, 2003, "Regulatory Network of Escherichia coli: Consistency between Literature Knowledge and Microarray Profiles", *Genome Research*, Vol. 13, pp. 2435-2443.

Herrgard, M. J., M. W. Covert and B. O. Palsonn, 2003, "Reconciling Gene Expression Data with Genome-Scale Regulatory Network Structures", *Genome Research*, Vol. 13, pp. 2423-2434.

Herrgard, M. J., M. W. Covert and B. O. Palsonn, 2004, "Reconstruction of Microbial Transcriptional Regulatory Networks", *Current Opinion in Biotechnology,* Vol. 15, pp. 70-77.

Harbison, C. T., D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, R. A. Young, 2004, "Transcriptional Regulatory Code of A Eukaryotic Genome", *Nature*, Vol. 431, pp. 99-104.

Hartemink, A. J., D. K. Gifford, T. S. Jaakkoia and R. A. Young, 2002, "Combining Location and Expression Data for Principled Discovery of Genetic Regulatory Networks", in *Pacific Symposium on Biocomputing* (Kauai, 2002).

Holloway, A. J., R. K. van Laar, R. W. Tothill and D. D. Bowtell, 2002, "Options Available -from Start to Finish- for Obtaining Data from DNA Microarrays II", *Nature Genetics*, Vol. 32 (Suppl), pp. 481-489.

Ihmels, J., G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv and N. Barkai, 2002, "Revealing Modular Organization in the Yeast Transcriptional Network", *Nature Genetics,* Vol. 31, pp. 370-377.

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabasi, 2000, "The Large-Scale Organization of Metabolic Networks", *Nature*, Vol. 407, pp. 651-654.

Klein, C. J., J. J. Rasmussen, B. Ronnow, L. Olsson and J. Nielsen, 1999, "Investigation of the Impact of MIG1 and MIG2 on the Physiology of *Saccharomyces cerevisiae*", *Journal of Biotechnology*, Vol. 68, pp. 197-212.

Kwast, K. E., L. C. Lai, N. Menda, D. T. James 3rd, S. Aref and P. V. Burke, 2002, "Genomic Analyses of Anaerobically Induced Genes in *Saccharomyces cerevisiae*: Functional Roles of Rox1 and Other Factors in Mediating the Anoxic Response", *Journal of Bacteriology*, Vol. 184, pp. 250-265.

Laub, M. T., S. L. Chen, L. Shapiro and H. H. McAdams, 2002, "Genes Directly Controlled by CtrA, a Master Regulator of the Caulobacter cell cycle", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 99, pp. 4632-4637.

Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford and R. A. Young, 2002, "Transcriptional Regulatory Networks in *Saccaromyces cerevisiae"*, *Science,* Vol. 298, pp. 799-804.

Liu X. S., D. L. Brutlag and J. S. Liu, 2002, "An Algorithm for Finding Protein DNA Binding Sites with Applications to Chromatin-Immunoprecipitation Microarray Experiments", *Nature Biotechnology*, Vol. 20, pp. 835-839.

Luscombe, N. M., M. M. Babu, H. Yu, M. Snyder, S. A. Teichmann and M. Gerstein, 2004, "Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes", *Nature,* Vol. 431, pp. 308-312.

Madan Babu, M. and S. A.Teichmann, 2003, "Evolution of Transcription Factors and the Gene Regulatory Network in *Escherichia coli*", *Nucleic Acids Research*, Vol. 31, pp. 1234-1244.

Matys, V., E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D. U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, E. Wingender, 2003, "TRANSFAC: Transcriptional Regulation, from Patterns to Profiles", *Nucleic Acids Research*, Vol. 31, pp. 374-378.

Milo, R., S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer and U. Alon, 2004, "Superfamilies of evolved and designed networks", *Science*, Vol. 303, pp. 1538-1542.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, 2002, "Network Motifs: Simple Building Blocks of Complex Networks", *Science*, Vol. 298, pp. 824-827.

MIPS (Munich Information Center for Protein Sequences), http://db.yeastgenome.org/, 2005.

Oltvai, Z. N. and A. L. Barabasi, 2002, "Systems Biology. Life's Complexity Pyramid", Science, Vol. 298, pp. 763-764.

Ogawa, N. and Y. Oshima, 1990, "Functional Domains of a Positive Regulatory Protein, PHO4, for Transcriptional Control of the Phosphatase Regulon in *Saccharomyces cerevisiae*", *Molecular and Cellular Biology*, Vol. 10, pp. 2224-2236.

Ostergaard, S., L. Olsson and J. Nielsen, 2000, "Metabolic Engineering of Saccharomyces cerevisiae", *Microbiology and Molecular Biology Reviews,* Vol. 64, No. 1, pp. 34-50.

Patil, K. R. and J. Nielsen, 2005, "Uncovering Transcriptional Regulation of Metabolism by Using Metabolic Network Topology", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 102, pp. 2685-2689.

Pe'er, D., A. Regev, G. Elidan and N. Friedman, 2001, "Inferring Subnetworks from Perturbed Expression Profiles", *Bioinformatics*, Vol. 17, Suppl 1, S215-S224.

Perez-Rueda, E. and J. Collado-Vides, 2000, "The Repertoire of DNA-Binding Transcriptional Regulators in *Escherichia coli* K-12", *Nucleic Acids Research,* Vol. 28, pp. 1838-1847.

Polish, J.A., J. H. Kim, and M. Johnston, 2004, "How the Rgt1 Transcription Factor of *S. cerevisiae* Is Regulated by Glucose", *Genetics,* (2004 Oct 16).

Ptashne, M. and A. Gann, 2002, "Genes and Signals". *Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press*.

Quackenbush, J., 2002, "Microarray Data Normalization and Transformation", *Nature Genetics*, Vol. 32 (Suppl), pp. 496-501.

Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett and E. Kanin, 2000, "Genome-Wide Location and Function of DNA Binding Proteins", *Science*, Vol. 290, pp. 2306-2309.

Rodrigues-Pousada, C. A., T. Nevitt, R. Menezes, D. Azevedo, J. Pereira and C. Amaral, 2004, "Yeast Activator Proteins and Stress Response: An Overview", *FEBS Letters*, Vol. 567, pp. 80-85.

Ronen, M., R. Rosenberg, B. I. Shraiman and U. Alon, 2002, "Assigning Numbers to the Arrows: Parameterizing a Gene Regulation Network by Using Accurate Expression Kinetics", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 99, pp. 10555-10560.

Salgado, H., A. Santos-Zavaleta, S. Gama-Castro, D. Millan-Zarate, E. Diaz-Peredo, F. Sanchez-Solano, E. Perez-Rueda, C. Bonavides-Martinez and J. Collado-Vides, 2001, "RegulonDB (version 3.2): Transcriptional Regulation and Operon Organization in *Esherichia coli* K-12", *Nucleic Acids Research*, Vol. 29, pp. 72-74.

Segal, E., M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller and N. Friedman, 2003, "Module Networks: Identifying Regulatory Modules and Their Condition-Specific Regulators from Gene Expression Data", *Nature Genetics*, Vol. 34, pp. 166-176.

Shen-Orr, S.S., R. Milo, S. Mangan and U. Alon, 2002, "Network Motifs in the Transcriptional Regulation Network of *Escherichia coli"*, *Nature Genetics,* Vol. 31, pp. 64-68.

Stitt, M. and A. R. Fernie, 2003, "From Measurements of Metabolites to Metabolomics: An 'On the Fly' Perspective Illustrated by Recent Studies of Carbon-Nitrogen Interactions", *Current Opinion in Biotechnology*, Vol. 14, pp. 136-144.

Struhl, K., 1999, "Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes", *Cell,* Vol. 98, pp. 1-4.

Svetlov, V. V. and T. G. Cooper, 1998, "The *Saccharomyces cerevisiae* GATA Factors Dal80p and Deh1p Can Form Homo- and Heterodimeric Complexes", *Journal of Bacteriology*, Vol. 180, pp. 5682-5688.

Tai, S. L., V. M. Boer, P. Daran-Lapujade, M. C. Walsh, J. H. de Winde, J. Daran and J. T. Pronk, 2005, "Two-Dimensional Transcriptome Analysis in Chemostat Cultures. Combinatorial Effects of Oxygen Availability and Macronutrient Limitation in *Saccharomyces cerevisiae*", *The Journal of Biological Chemistry*, Vol. 280, pp. 437-447.

Tanay, A. and R. Shamir, 2001, "Computational Expansion of Genetic Networks", *Bioinformatics*, Vol. 17, Suppl 1, S270-S278.

Teichmann, S. A. And M. M. Babu, 2004, "Gene Regulatory Network Growth by Duplication", *Nature Genetics*, Vol. 36, pp. 492-496.

TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins, http://www.gene-regulation.com/, 2005.

Wang, W., J. M. Cherry, D. Botstein, and H. Li, 2002, "A Systematic Approach to Reconstructing Transcription Networks in *Saccharomyces cerevisiae"*, *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 99, pp. 16893-16898.

Wingender, E., X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, R. Ohnhauser, M. Pruss, F. Schacherer, S. Thiele, S. Urbach S, 2001, "The TRANSFAC System on Gene Expression Regulation", *Nucleic Acids Research*, Vol. 29, pp. 281-283.

Wyrick, J.J and R. A. Young, 2002, "Deciphering Gene Expression Regulatory Networks", *Current Opinion in Genetics & Development,* Vol. 12, pp. 130-136.

"Yeast Search for Transcriptional Regulators And Consensus Tracking (YEASTRACT)", [February 10, 2005] http://www.yeastract.com, [November 19, 2005].

Yu, H., N. M. Luscombe, J. Quian and M. Gerstein, 2003, "Genomic Analysis of Gene Expression Relationships in Transcriptional Regulatory Networks", *Trends in Genetics*, Vol. 19, pp. 422-427.

Zeitlinger, J., I. Simon, C. T. Harbison, N. M. Hannett, T. L. Volkert, G. R. Fink and R. A. Young, 2003, "Program-Specific Distribution of a Transcription Factor dependent on partner Transcription Factor and MAPK Signaling", *Cell*, Vol. 113, pp. 395-404.

Zhu, J. and M. Q. Zhang, 1999, "SCPD: A Promoter Database of the Yeast Saccharomyces cerevisiae", *Bioinformatics,* Vol. 15, pp. 607-611.

APPENDIX A: COMPUTER CODE FOR TOPOLOGICAL CALCULATIONS

```
PROGRAM TOPOLOGY
INTEGER S1(4,4),S2(4,4),S3(4,4),
&  S4(4,4),S5(4,4),S6(4,4),
& S7(4,4), S8(4,4),S9(4,4),
& S10(4,4),S11(4,4),S12(4,4),
& S13(4,4),S14(4,4),
& S15(4,4),S16(4,4)
& ,SPL(4,4),MAX
REAL TOP,TOP2,NUM
DATA  S1/16*0/,S2/16*0/,S3/16*0/,
& S4/16*0/,S5/16*0/,S6/16*0/,
& S7/16*0/,S8/16*0/,S9/16*0/
& ,S10/16*0/,S11/16*0/,S12/16*0/,
& S13/16*0/,S14/16*0/,S15/16*0/
& ,S16/16*0/
DATA SPL/16*0/
20      FORMAT (I5)
30      FORMAT (I8)
OPEN (3, FILE='S1.txt',STATUS='OLD')
OPEN (5, FILE='S2.txt',STATUS='OLD')
OPEN (7, FILE='S3.txt',STATUS='OLD')
OPEN (11, FILE='S4.txt',STATUS='OLD')
OPEN (13, FILE='S5.txt',STATUS='OLD')
OPEN (15, FILE='S6.txt',STATUS='OLD')
OPEN (17, FILE='S7.txt',STATUS='OLD')
OPEN (19, FILE='S8.txt',STATUS='OLD')
OPEN (21, FILE='S9.txt',STATUS='OLD')
OPEN (23, FILE='S10.txt',STATUS='OLD')
OPEN (25, FILE='S11.txt',STATUS='OLD')
```

```
OPEN (27, FILE='S12.txt',STATUS='OLD')
OPEN (29, FILE='S13.txt',STATUS='OLD')
OPEN (31, FILE='S14.txt',STATUS='OLD')
OPEN (33, FILE='S15.txt',STATUS='OLD')
OPEN (35, FILE='S16.txt',STATUS='OLD')
OPEN (9, FILE='SPL.txt',STATUS='NEW')
DO I=1,4
DO J=1,4
READ (3,10) S1(J,I)
READ (5,10) S2(J,I)
READ (7,10) S3(J,I)
READ (11,10) S4(J,I)
READ (13,10) S5(J,I)
READ (15,10) S6(J,I)
READ (17,10) S7(J,I)
READ (19,10) S8(J,I)
READ (21,10) S9(J,I)
READ (23,10) S10(J,I)
READ (25,10) S11(J,I)
READ (27,10) S12(J,I)
READ (29,10) S13(J,I)
READ (31,10) S14(J,I)
READ (33,10) S15(J,I)
READ (35,10) S16(J,I)
END DO
END DO
MAX=0
TOP=0.
NUM=0.
TOP2=0.
DO I=1,4
DO J=1,4
IF ((S15(J,I).EQ.0).AND.(S16(J,I).EQ.1)) THEN
```

```
SPL(J,I)=16
END IF
IF ((S14(J,I).EQ.0).AND.(S15(J,I).EQ.1)) THEN
SPL(J,I)=15
END IF
IF ((S13(J,I).EQ.0).AND.(S14(J,I).EQ.1)) THEN
SPL(J,I)=14
END IF
IF ((S12(J,I).EQ.0).AND.(S13(J,I).EQ.1)) THEN
SPL(J,I)=13
END IF
IF ((S11(J,I).EQ.0).AND.(S12(J,I).EQ.1)) THEN
SPL(J,I)=12
END IF
IF ((S10(J,I).EQ.0).AND.(S11(J,I).EQ.1)) THEN
SPL(J,I)=11
END IF
IF ((S9(J,I).EQ.0).AND.(S10(J,I).EQ.1)) THEN
SPL(J,I)=10
END IF
IF ((S8(J,I).EQ.0).AND.(S9(J,I).EQ.1)) THEN
SPL(J,I)=9
END IF
IF ((S6(J,I).EQ.0).AND.(S7(J,I).EQ.1)) THEN
SPL(J,I)=7
END IF
IF ((S5(J,I).EQ.0).AND.(S6(J,I).EQ.1)) THEN
SPL(J,I)=6
END IF
WRITE (9,20)   SPL(J,I)
TOP=TOP+SPL(J,I)
IF (SPL(J,I).NE.0) THEN
TOP2=TOP2+SPL(J,I)
```

```
NUM=NUM+1.
END IF
IF (SPL(J,I).GT.MAX) THEN
MAX=SPL(J,I)
END IF
END DO
PRINT*, 'AVERAGE WITHOUT ZEROS =   ', TOP2/NUM
PRINT*, 'AVERAGE WITH ZEROS =    ', TOP/16.0
PRINT*, 'DIAMETER=   ',MAX
END
```

APPENDIX B: COMPUTER CODE FOR *Z* SCORE CALCULATIONS

```
PROGRAM SZS
CHARACTER * 7 GEN(7000),OKU(1000),DOS,DOSYA(200),NAME
REAL DEGER(7000)
INTEGER DS,FILEN,GENSAY,OKUSAY,FILEN2
C 5      format (I7)
10       FORMAT (A7)
20       FORMAT (F13.5)
30       FORMAT (I7)
40       FORMAT (A7,F13.5)
42       FORMAT (A9,F13.5)
43       FORMAT(A7,A12)
FILEN=10
PRINT *, "
PRINT *, "
PRINT *, "
PRINT *, ' '
PRINT *, ' '
PRINT *, ' '
PRINT *, ' '
C      PART 1: NAME READING
PRINT *, '>>>>>> THE NAME
PRINT *,'*** IMPORTANT NOTICE
PRINT *, ' '
READ *, DOS
PRINT *, ' '
PRINT *,'>>>>>> THE QUESTION
PRINT *,'   PRESS
PRINT *, ' '
READ *, DS
```

```
DO I5=1,DS
PRINT *,I6, ' NAME OF THE FILE
READ *, DOSYA(I2)
PRINT *, ' '
END DO
C      PART 2: DATA READING
OPEN (2, FILE='sonuclar.txt',STATUS='NEW')
OPEN (3, FILE=DOS//'.txt',STATUS='OLD')
READ (3,30) GENSAY
DO I6=1,GENSAY
READ (3,10) GEN (I4)
END DO
DO I5=1,GENSAY
READ (3,20) DEGER(I4)
END DO
PRINT *, ' '
PRINT *, ' ****',DOS,'. ***'
PRINT *, ' '
C       PART 3: START OF OPERATIONS
DO I5=1,DS
PRINT *, ' '
PRINT *,'***',  DS, '.  ***'
FILEN=FILEN+1
WRITE (NAME,10) DOSYA(I4)
OPEN (FILEN, FILE=NAME//'.txt',STATUS='OLD')
READ (FILEN,30) OKUSAY
DO I5=1,OKUSAY
READ (FILEN,11) OKU (I6)
END DO
FILEN2=100-FILEN
OPEN (FILEN2,FILE=NAME//'sonuc.txt',STATUS='NEW')
SUM=0.0
DO I5=1,OKUSAY
```

```
DO I6=1,GENSAY
IF (GEN(I9).EQ.OKU(I3)) THEN
SUM=SUM+DEGER(I2)
WRITE (FILEN2,40) OKU(I8),DEGER(I6)
print *,  OKU(I4),DEGER(I5)
GOTO 50
END IF
END DO
WRITE (FILEN2,43) OKU(I6), '  CAN NOT BE FOUND'
PRINT*, OKU(I6), '  CAN NOT BE FOUND'
50      continue
END DO
WRITE (2,10) SUM
WRITE (FILEN2,13) '       '
WRITE (FILEN4,28) 'TOTAL = ', SUM
PRINT *, ' '
PRINT *,DS,'. TOTAL=', SUM
END DO
PRINT *, ' '
PRINT *, ' PROGRAM COMPLETED '
```

APPENDIX C: COMPUTER CODE FOR NEW *Z* SCORE CALCULATIONS

```
PROGRAM RANDOM
CHARACTER * 5 DOS
REAL  DEGER (6500),RAND,ra,SUM(10000),TOPSUM,AVER,STD,yeniz,GEN
REAL zscore
INTEGER IRA
12       FORMAT (F15.2)
C 15     FORMAT (I3)
PRINT *, FILE NAME
PRINT *,' *** IMPORTANT NOTICE
READ *, DOS
OPEN (5, FILE=DOS//'.txt',STATUS='OLD')
C        OPEN (4, FILE='SONUC.txt',STATUS='NEW')
C        OPEN (5, FILE='random.txt',STATUS='NEW')
OPEN (7, FILE='yeni_z_score.txt',STATUS='NEW')
DO I=1,6400
READ (2,10) DEGER (I)
END DO
dummy=100
100     CONTINUE
PRINT *, NUMBER OF GENES
READ *, GEN
TOPSUM=0.0
DO K=1,10000
SUM(K)=0.0
DO J=1,GEN
ra = 6400.0*rand(0)
IRA=NINT (RA)
C        WRITE (5,11) IRA
```

```
SUM(K)=SUM(K) + 1.0/SQRT(GEN)*DEGER(IRA)
END DO
C       WRITE (5,23) SUM (K)
TOPSUM=SUM(K)+TOPSUM
END DO
AVER=TOPSUM/10000.0
VARIAN=0.0
DO I5=1,10000
VARIAN= (SUM (I3) - AVER)**2.0 + VARIAN
END DO
STD= SQRT(VARIAN /9999)
PRINT *, Z SCORE?
READ *, zscore
yeniz=(zscore-Aver)/STD
WRITE (7,12) yeniz
PRINT *, 'Yeni Z-Score =' , yeniz
PRINT *, " "
PRINT *, "
PRINT *, " "
If (dummy.EQ.100) then
GOTO 60
end if
stop
end
```

APPENDIX D: HIGHEST SCORING TFS BEFORE CORRECTION

Table D.1. Oxygen-responsive top 10 TFs before correction. Under carbon, nitrogen, phosphorous and sulfur limitation regimes. Hap2p, Hap3p and Hap4p and Hap5p were counted as one TF because they form a complex and their scores are very close to each other

ANA-C-AER-C		ANA-N-AER-N		ANA-P-AER-P		ANA-S-AER-S	
16,12	Sok2p	17,21	Upc2p	17,44	Hap2p	19,16	Hap5p
15,39	Msn4p	16,28	Hap5p	17,43	Hap3p	18,94	Hap2p
15,00	Yap1p	16,20	Hap2p	17,41	Hap5p	18,92	Hap4p
14,18	Gcn4p	16,19	Hap4p	17,37	Hap4p	18,86	Hap3p
14,10	Rap1p	16,13	Hap3p	15,56	Rap1p	18,48	Upc2p
13,96	Cat8p	15,37	Sok2p	15,00	Yap1p	18,40	Yap1p
13,89	Hap2p	14,32	Yap1p	14,95	Sok2p	17,02	Cad1p
13,88	Hap3p	13,17	Rox1p	14,63	Upc2p	16,67	Sok2p
13,86	Hap4p	13,14	Cad1p	14,10	Rox1p	15,58	Rox1p
13,82	Hap5p	12,03	Yap5p	13,92	Msn4p	15,55	Msn4p
12,13	Yap5p	12,02	Hcm1p	13,89	Yap5p	15,46	Hcm1p
12,00	Hcm1p	11,64	Msn4p	12,90	Tos8p	15,07	Yap5p
11,90	Stp1p	11,51	Stp1p	12,84	Gcn4p	14,94	Swi4p

Table D.2. Top 10 TFs before correction between nutrient limitations (aerobic). Between carbon, nitrogen, phosphorous and sulfur limitation regimes

AER-C-N		AER-C-P		AER-C-S		AER-N-P		AER-N-S		AER-P-S	
18,56	Sok2p	16,49	Sok2p	19,25	Gcn4p	14,04	Sok2p	17,86	Yap1p	15,29	Msn4p
16,92	Rap1p	15,61	Msn4p	19,04	Sok2p	12,99	Gcn4p	17,63	Gcn4p	14,39	Yap1p
15,66	Yap1p	14,89	Rap1p	18,65	Yap1p	12,96	Rap1p	17,59	Sok2p	14,29	Gcn4p
15,48	Msn4p	13,61	Yap1p	17,68	Msn4p	12,34	Stp1p	15,98	Rap1p	12,70	Sok2p
15,30	Gcn4p	13,02	Gcn4p	17,53	Cad1p	11,42	Stp2p	15,23	Cad1p	12,21	Pdr3p
14,69	Cad1p	12,76	Stp2p	15,75	Rap1p	11,37	Yap1p	14,53	Msn4p	12,15	Stp1p
13,92	Hcm1p	12,65	Stp1p	15,48	Hcm1p	11,27	Cad1p	14,50	Pdr3p	11,78	Cad1p
13,72	Pdr1p	12,45	Pdr3p	14,99	Cin5p	11,20	Ste12p	13,79	Cin5p	11,42	Stp2p
13,45	Mcm1p	12,17	Hcm1p	14,83	Pdr3p	11,10	Gln3p	13,75	Stp1p	10,81	Pdr1p
13,18	Tos8p	11,87	Cad1p	14,17	Cat8p	10,33	Msn4p	13,46	Swi4p	10,71	Rap1p

Table D.3. Top 10 TFs before correction between nutrient limitations (anaerobic). Between carbon, nitrogen, phosphorous and sulfur limitation regimes. Hap2p, Hap3p and Hap4p and Hap5p were counted as one TF because they form a complex and their scores are very close to each other

ANA-C-N		ANA-C-P		ANA-C-S		ANA-N-P		ANA-N-S		ANA-P-S	
15,98	Rap1p	15,13	Rap1p	19,77	Rap1p	14,97	Gcn4p	14,14	Gcn4p	13,79	Gcn4p
15,93	Sok2p	14,08	Sok2p	18,42	Sok2p	14,96	Sok2p	12,99	Sok2p	12,38	Sok2p
13,80	Yap1p	13,37	Yap1p	18,08	Gcn4p	13,23	Msn4p	12,21	Msn4p	12,24	Yap1p
13,30	Gcn4p	12,37	Gcn4p	15,96	Cad1p	12,57	Stp1p	12,13	Yap1p	11,35	Cad1p
12,75	Msn4p	12,18	Hap5p	15,75	Yap5p	12,50	Stp2p	11,85	Cad1p	11,17	Rap1p
12,61	Cad1p	12,17	Hap4p	15,55	Hap4p	12,33	Yap1p	10,89	Stp1p	11,13	Stp2p
12,57	Yap5p	12,17	Hap2p	15,48	Hap5p	12,05	Rap1p	10,26	Stp2p	11,07	Msn4p
12,23	Hap4p	12,10	Hap3p	15,43	Swi4p	11,53	Cad1p	9,98	Pdr3p	10,93	Stp1p
12,21	Hap5p	11,85	Msn4p	15,43	Hap2p	11,40	Gln3p	9,71	Met4p	10,30	Pdr1p
12,15	Pdr3p	11,15	Pdr1p	15,41	Hap3p	11,38	Skn7p	9,71	Met28p	10,30	Met4p
12,12	Hap2p	11,07	Stp2p	15,17	Yap1p	14,97	-	-	-	-	-
12,12	Hap3p	11,07	Stp1p	14,81	Msn4p	14,96	-	-	-	-	-