INTEGRATED ANALYSIS OF METABOLOME PROFILES AND GENE EXPRESSION IN RESPIRATORY DEFICIENT DELETION MUTANTS OF *SACCHAROMYCES CEREVISIAE*

by

Pınar Pir

B.S., Chemical Engineering, Boğaziçi University, 1998

M.S., Chemical Engineering, Boğaziçi University, 2001

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

Graduate Program in Chemical Engineering

Boğaziçi University

2005

ACKNOWLEDGEMENTS

I would like to thank my advisors Prof. Kutlu Ö. Ülgen and Prof. Z. İlsen Önsan for their support and help throughout my studies. I would like to express my gratitude to Prof. Betül Kırdar and Prof. Stephen G. Oliver for their help and encouragement besides making this project possible. I would like to thank to Prof. Serpil Takaç for accepting to be a member of my jury and critical reading of my thesis.

This work was supported by Boğaziçi University Research Fund through projects 03S108, 03A504, 05HA502, 04HA503D, by DPT-03K120250, and by grants from the Wellcome Trust and the COGEME. The scholarship provided by TUBITAK-BAYG is gratefully acknowledged.

I would like to thank to all academic, administrative and technical staff of B.Ü. Department of Chemical Engineering for their help and support throughout my education. Also I would like to thank to staff and students of U.M. Faculty of Life Sciences for their help and support during my studies in University of Manchester.

I would like to thank to my parents for their endless support and patience. I am lucky to have many brothers and sisters: Thanks to Çağlar for being the Perfect Brother, thanks for existing at all... Thanks to my sisters; Diğdem, Duygu, Hilal and Zorana, I believe they are angels from heaven... Thanks to Dandik for the short but the great accompany... Being a “group” is beyond being “collegues”: Thanks to all GenetikGeyikler and Yeasties for their friendship and support... Thanks to Tunahan and Yalçın for the endless brain stormings and thanks to Deniz for sharing the “tips of life”... Thanks to Mete, Çiğdem, Barış, Burcu and Ahmet for their everlasting friendship... I have sweet homes and beloved cities, full of friends and joy: Thanks to İstanbul, Manchester, Çanakkale, Ankara, Delft and Copenhagen... I have milestones in my life: thanks to Burak for the questions and thanks to Sedat for the answers...

Last, but not the least, thanks to Time. It is the “teacher” with no mercy and it is the cure for “anything”... Thanks to Time for passing by... Now I am older and wiser...

ABSTRACT

A study on transcriptional and metabolic response of *S. cerevisiae* cells to environmental and genetic perturbations was made. Deletion mutants with BY4743 backgound were used. Homozygous deletion mutants *hoΔ/hoΔ* (reference strain), *hap4Δ/hap4Δ*, *oxa1Δ/oxa1Δ*, *bcs1Δ/bcs1Δ*, *rip1Δ/rip1Δ*, *mig1Δ/mig1Δ*, and *mba1Δ/mba1Δ* and heterozygous deletion mutants *HAP4*/*hap4Δ*, *RIP1*/*rip1Δ* and *RIP1*/*mig1Δ* were grown aerobically in continuous chemostat reactors. The samples were collected after three retension times and analysed for biomass, mRNA and metabolite contents.

Transcriptional response of the yeast cells to gene deletions and growth conditions enabled investigation of transcriptional regulation of central carbon metabolism. Novel regulation information was extracted using present data set.

Integration of metabolic and transcriptome data was accomplished using partial least squares (PLS) method, which enables modelling of metabolic data using transcriptome data. Results of PLS indicate the genes which may mediate the changes in metabolites due to perturbations applied.

Functional annotation to 17 unknown ORFs were annotated functions depending on the genes they are transcriptional correlated. A network among the genes with correlated transcription was constructed and its prediction power of existing annotations was tested. Data from other sources (amino-acid sequence similarity, protein-protein interaction, shared transcription factors) were combined with transcription data to improve the network and prediction power. It was concluded that integration of information from various sources enables better prediction of functions of known genes and allows more unknown ORFs to be annotated.

Results of this study revealed transcriptional regulation relations unknown previously, allowed functional annotation of more than 500 unknown ORFs and a link between metabolome and transcriptome profiles of *S. cerevisiae* is constructed.

ÖZET

S. cerevisiae hücrelerinin çevresel ve genetik değişiklere verdiği transkripsiyon tepkisi and metabolik tepkiler üzerine bir çalışma yapılmıştır. BY4743 temelli homozigot delesyon mutantları *hoΔ/hoΔ* (referans alınan suş), *hap4Δ/hap4Δ*, *oxa1Δ/oxa1Δ*, *bcs1Δ/bcs1Δ*, *rip1Δ/rip1Δ*, *mig1Δ/mig1Δ* ve heterozigot delesyon mutantları *HAP4/hap4Δ*, *RIP1/rip1Δ*, *MIG1/mig1Δ* havalandırmalı ve sürekli chemostat reaktörlerde büyütülmüştür. Üç retensiyon zamanı sonrasında toplanan örnekler, biyokütle, mRNA ve metabolit konsantrasyonu analizlerine tabi tutulmuştur.

Maya hücrelerinin gen delesyonlarına ve büyüme koşullarına verdiği transkripsiyon tepkisi, merkezi karbon metabolizmasının transkripsiyon regulasyonunun incelenmesini mümkün kılmıştır. Elde edilen verilerle, yeni regulasyon bilgilerine ulaşılmıştır.

Metabolit ve transkripsiyon verilerinin bütünleştirilmesi, kısmı en küçük kareler (PLS) metodu ile gerçekleştirilmiştir. Böylece transkripsiyon verileri kullanarak metabolit miktarlarının modellenmesi mümkün olmuştur. PLS sonuçları, büyüme koşullarında yapılan değişiklikler sonucunda metabolit miktarlarının değişmesine sebep olan genlerin belirlenmesini sağlamıştır.

Transkripsiyonu korelasyon gösteren genler arasında bir ağ oluşturulmuş ve bu ağın varolan fonksiyon atamalarını tahmin edebilme gücü test edilmiştir. Ağın zenginleştirilmesi ve tahmin gücünün arttırılması amacı ile diğer veri kaynaklarından alınan bilgiler (aminoasit dizilimi benzerliği, protein-protein etkileşimi, paylaşılan transkripsiyon faktörleri) transkripsiyon verileriyle bütünleştirilmiştir. Çeşitli bilgi kaynaklarından alınan bilgilerin bütünleştirilmesinin, bilinen genlerin fonsiyonlarının daha iyi tahmin edilebilmesini ve fonksiyonu bilinmeyen birçok gene fonksiyon atanmasını sağlamaktadır.

Bu çalışmanın sonuçları, daha önce bilinmeyen regulasyon ilişkilerinin ortaya çıkarılmasını, bir çok bilinmeyen gene fonksiyon atanmasını ve metabolitler ile transkripsiyon profilleri arasında bir ilişki kurulmasını sağlamıştır.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

a	Number of levels of factor τ
a	Mating type
b	Number of levels of factor β
b	Inner relation
B	Regression vector
c	Number of levels of factor γ
D	Dilution rate (hr^{-1})
e	Residual matrix
E	Residual matrix
Fo	F-distribution
G	ORFs annotated to a specific term.
i	Index for levels of factor τ
j	Index for levels of factor β
k	Index for levels of factor γ
l	Index for replicates
n	Number of replicates
n	ORFs in the list investigated
N	ORFs annotated to all terms
p	P-value
p	Loading matrix of X
q	Loading matrix of Y
t	Score matrix of X
t	Transpose of a matrix
u	Score matrix of Y
$\bar{u}$	Estimated score matrix of Y
x	ORFs in the list annotated to the specific term
X	Transcriptome matrix
y	Output variable
Y	Metabolic data matrix

SS_A	Sum of squares for factor A
MS_A	Mean sum of squares for factor A
α	Mating type
β	Factor in linear modeling
γ	Factor in linear modeling
Δ	Gene deletion
ε	Random error in linear modelling
μ	Mean of the outputs
ρ	Pearson correlation coefficient
τ	Factor in linear modeling
ϕ	Number of ORFs annotated to a term divided by number of all ORFs
AAA1	Gene name
AAA1/aaa1Δ	Heterozygous deletion mutant of gene *AAA1*
Aaa1p	Protein product of gene *AAA1*
aaa1Δ/aaa1Δ	Homozygous deletion mutant of gene *AAA1*
AcCoA	Acetyl coenzyme A
ADH	Alcohol dehydrogenase
ADP	Adenosine diphosphate
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
AVID	Annotation Via Integration of Data
BIND	The Biomolecular Interaction Network
BMBF	Federal Ministry of Education and Research
bp	Base pair
BY4743	Cell line
cDNA	Complementary DNA
CE	Capillary electrophoresis
CoA	Coenzyme A
cRNA	Complementary RNA
Cy3	Fluorescent dye (green)

Cy5	Fluorescent dye (red)
CYGD	Comprehensive Yeast Genome Database
DAHP	dihydroxy acetone phosphate
DC	Downregulation at glucose limitation
DEPC	Diethylpyrocarbonate
DH	Upregulation at hap4 deletion
DN	Downregulation at ammonium limitation
DNA	Deoxyribonucleic acid
DO	Dissolved oxygen
DR1	Downregulation at D = $0.1hr^{-1}$
DR2	Downregulation at D = $0.2hr^{-1}$
dUTP	Deoxyuridine triphosphate
EDTA	Ethylene diamine tetra acetic acid
ES	Electron spray
EUROSCARF	European *Saccharomyces Cerevisiae* Archive for Functional Analysis
FAD	Flavin adenine dinucleotide
FTIR	Fourier Transform Infrared
GAL	Galactose
GC	Gas chromatography
GeneFAS	Gene Function Annotation System
GO	Gene Ontology
GRID	General Repository of Interaction Databases
HEPES	4-2-hydroxyethyl-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
IPA	Isopropyl alcohol
IR	Infrared spectroscopy
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LIF	Laser-induced fluorescence
LV	Latent variable
MALDI-TOF	Matrix-assisted laser desorption/ionization – time of flight
MIPS	Munich Information Center on Protein Sequences
mRNA	Messenger RNA

MS	Mass spectrometry
mTn	Mini transposon
NAD	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
NTP	Nucletide triphosphate
OAA	Oxaloacetate
ORF	Open reading frame
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDA	Photodiode array
PDC	Pyruvate decarboxylase
PFK	Phosphofructokinase
PID	Proportional-Integral-Derivative
PLS	Partial Least Square
PPI	Protein-protein interaction
PredPower	Prediction power
PRESS	Prediction sum of squares
PYC	Pyruvate kinase
RNA	Ribonucleic acid
SGD	Saccharomyces Genome Database
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TAE	Tris-Borate-EDTA
TBE	Tris-Acetate-EDTA
TCA	Tricarboxylic acid
TF	Transcription factor
TLC	Thin-layer chromatography
UC	Upregulation at carbon limitation
UN	Upregulation at ammonium limitation
UR1	Upregulation at D = $0.1hr^{-1}$
UR2	Upregulation at D = $0.2hr^{-1}$
UV	Ultraviolet
YEASTRACT	Yeast Search for Transcriptional Regulators And Consensus Tracking

YPA-D	Yeast extract – Peptone – Agar – D-glucose
YPA-E	Yeast extract – Peptone – Agar – Ethanol
YPD	Yeast extract – Peptone – D-glucose
YPD	Yeast Proteome Database

1. INTRODUCTION

Genetic research on *Saccharomyces cerevisiae* has been crucial because of industrial importance of this unicellular eukaryotic yeast. Its ease of genetic manipulation and similarity to higher eukaryotes placed *S. cerevisiae* as a model organism in functional genomics and systems biology.

Genome sequencing of *S. cerevisiae* genome was completed in 1996 (Goffeau *et al.*, 1996) and a library of deletion mutants was constructed covering deletion mutants of almost all open reading frames (ORFs) (EUROSCARF, 2003). The literature survey on the yeast *S. cerevisiae* and the recent genetic research focused on *S. cerevisiae* is given in the "Literature Survey" section.

In this study, transcriptome and metabolome data were collected from *S. cerevisiae* deletion mutants growing aerobically in nutrient limited chemostats. The experiments conducted are described and the list of the materials used are given in "Experimental Materials and Methods" under the section "Materials and Methods". The software and mathematical methods used in data analysis are given in the "Mathematical Materials and Methods" subsection.

The results of this study relate to four main topics. First, transcriptional regulation of central carbon metabolism of *S. cerevisiae* was investigated along with the regulation of respiratory chain. Results are given in the section "Transcriptional Regulation of Central Carbon Metabolism of *S. cerevisiae*".

Integration of metabolome data and transcriptome data was made to identify the genes related to changes in metabolism data. The discussion of the methodology and the results are given in the section "Integration of Metabolome and Transcriptome Data".

The ORFs significantly affected by the perturbations were selected using linear modelling and ANOVA. 50 unknown ORFs reported to be affected in more than two perturbation experiments (including the experiments conducted in the present study) were

selected for detailed analysis. 17 of these ORFs were annotated with novel functions, a literature survey was carried out to find evidence to support these novel annotations. These annotations were compared with annotations made by automated annotation algorithms (AVID, GeneFas, String). The results and discussion of this work are given in the section, "Functional Annotation of Unknown Yeast ORFs".

Biological networks connecting the ORFs that were observed to have a shared property (co-expression, amino-acid sequence similarity, protein-protein interaction, shared transcription factors) were constructed and the prediction power of these networks for functional annotation of unknown ORFs were tested on GO biological process and molecular function term networks. More than 500 unknown ORFs were annotated novel functions. An algorithm was proposed and the software "PredPower" was developed which is composed of a group of MATLAB programs to carry out the computations. The algorithm and results are given in the section "Investigation of Prediction Power of Networks for Functional Annotation".

In this study, analysis of the collected experimental data via various bioinformatics tools elucidated transcriptional regulation relations that were unknown previously, allowed functional annotation of many unknown ORFs and a link between metabolome and transcriptome profiles of *S. cerevisiae* to be constructed. An overall picture of the biological facts elucidated in this study and further directions are given in section "Conclusions and Recommendations". This work provides further evidence that transcriptome and metabolome data are invaluable resources in functional genomics and systems biology research.

2. LITERATURE SURVEY

2.1. History and Importance of Yeast Research

Yeasts are unicellular fungi which have been used by mankind to produce alcoholic beverages and leaven bread dough for millenia. The brewing of beer probably represented the world's first biotechnology and yeast played a central role in food production and conservation due to its ability to ferment glucose to ethanol and carbon dioxide. Beside being useful in daily brewing and baking practice, yeast, (as a simple, unicellular eukaryote) developed into a powerful model system for biological research. Its prominent useful features are cheap and easy cultivation, short generation times, detailed genetic and biochemical knowledge accumulated in many years of research and ease of the application of molecular techniques for its genetic manipulation. Thus, this organism provides a highly suitable system to study biological processes that are relevant for many higher eukaryotes including man (Table 2.1).

After determination of the complete genome sequences of two prokaryotes, *M. genitalium* and *H. influenzae*, the first complete genomic DNA sequence of a eukaryote, *S. cerevisiae,* has been unraveled as a result of a worldwide scientific collaboration (Goffeau *et al.*, 1996). Bioinformatics has helped to assemble the DNA sequence of the 16 chromosomes in yeast and to extract the aminoacid sequences of more than 6000 open reading frames (ORFs) with a length of more than 99 codons. The analysis of the mitochondrial genome defined 32 ORFs, 4 of which are smaller than 100 amino acids.

2.2. Yeast Cytology and Mitochondrion

The yeast *S. cerevisiae* is a unicellular eukaryote. It has the major organelles nucleus, mitochondria, Golgi apparatus, secretory vesicles, endoplasmic reticulum, vacuoles and microbodies. The cytoplasm contains ribosomes and occasionally plasmids, and the structural organisation is maintained by a cytoskeleton. The cellular contents are encased by an envelope comprising plasma membrane, periplasm and cell wall. Yeast cells share many of the structural (and functional) features of higher eukaryotes, including the

mitochondria and, consequently, yeasts have been considered models for eukaryotic cell science. The presence of a cell wall is the major structural difference between a yeast cell and an animal cell (Walker, 1998).

Table 2.1. Milestones in the study and exploitation of yeast (Walker, 1998).

2000-6000 BC	Brewing (Sumeria, Babylonia), grape cultivation (Georgia), dough leavening (Egypt)
1680	Description of the microscopic appearance of yeasts (van Leeuwenhoek)
1830s	Alcoholic fermentation associated with yeast budding (Cagnaird-Latour in 1835)
	The name Saccharomyces cerevisiae, created for yeast observed in malt (Meyen in 1837)
	Sugar acting as food source for yeast growth (Schwann and Kützing in 1839)
Late 1800s	'All fermentations are correlative with physiological phenomena' and fermentations associated with yeast metabolism (Pasteur in 1857). The term 'enzyme' (Greek; 'in yeast') introduced by Kuhne in 1877. Isolation of single yeast cells and employment of pure yeast strains for brewing (Hansen, 1880-1883). The production of alcohol and CO_2 from sugar by cell-free yeast extracts (E. and H. Buchner). Birth of biochemistry as a new scientific discipline.
1920	Knowledge of yeast physiology, sexuality and phylogeny reviewed. (Guilliermond in 1920)
1930s-1940s	Genetic studies commenced on brewing yeast (Winge, Lausen, Lindegren).
	Demonstration of sexual reproduction and mating type system
1930-1960	Yeast taxonomy studies by Kluyver and the Delft School.
1970s	First transformation of yeast in 1978 (Hinnen, Hicks, Fink and Beggs)
1980s-1990s	First commercial pharmaceutical (hepatitis B vaccine) from recombinant yeast.
	Completion of *S. cerevisiae* genome project (1996).

S. cerevisiae cells are generally ellipsoidal in shape with a large diameter ranging from 5-10μm, and a small diameter of 1-7 μm. The mean cell volumes for a haploid and diploid cell are 29 and 55 μm^3, respectively. The mean cell size of *S. cerevisiae* also increases with age and the nutritional environment also affects cell morphology. The distinguishable cell types of *S. cerevisiae* are **a**, α, and **a**/α cells. The **a** and α haploid cells are able to undergo mating with nuclear fusion which results in creation of **a**/α diploid (Walker, 1998).

Yeast mitochondria structurally resemble those organelles found in higher eukaryotes. Mitochondria, are membrane-bound organelles that convert energy to forms

that can be used by metabolic reactions. Each mitochondrion is bounded by membranes creating two separate mitochondrial compartments: The internal matrix space and intermembrane space. The intermembrane space is chemically equivalent to the cytosol with respect to the small molecules it contains, the matrix space contains a highly selected set of small molecules because of the impermeable inner membrane. Each membrane contains a unique collection of proteins (Alberts *et al.*, 1994).

The major working part of the mitochondrion is the matrix space and the inner membrane that surrounds it. The inner membrane contains variety of transport proteins that make it selectively permeable to those small molecules that are metabolized or required by the many mitochondrial enzymes concentrated in the matrix space. The matrix enzymes include those that metabolize pyruvates and fatty acids to produce acetyl CoA and those that oxidize acetyl CoA in the citric acid cycle. The enzymes of the respiratory chain are embedded in the inner mitochondrial membrane, and they are essential for oxidative phosphorylation, which generates most of the yeast cell's ATP (Alberts *et al.*, 1994).

Most mitochondrial proteins of yeast (approximately 95 per cent) are encoded by nuclear genes, synthesized on free cytoplasmic polysomes, and finally imported into mitochondria. The circular mitochondrial DNA codes for only a few polypeptides: subunits I, II and III of cytochrome c oxidase; subunits 6 and 9 of the ATPase complex, apocytochrome b and some maturases required for intron removal. In addition to these, mitochondria have their own translational machinery. The components of the mitochondrial protein synthesis apparatus coded by mitochondrial genes are the two rRNAs (15S and 21S, constituents of the small and large ribosomal subunits), a complete set of tRNAs and a few mRNA species (Alberts *et al.*, 1994).

The isolation and characterization of mitochondrial mutants of *S. cerevisiae* has provided fundamental biochemical, genetic and molecular biological information on the biogenesis and function of mitochondria. Moreover, respiratory-deficient mutants of yeasts have been studied with regard to their possible exploitation in fermentation biotechnology. For example, the use of petite mutants of *S. cerevisiae* in alcohol fermentation has been suggested since alcohol yields can be quite high due to the cells' insensitivity to oxygen and their minimal diversion of carbon to cell growth. However, respiratory-deficient

petites grow more slowly than wild-type 'grande' strains of *S. cerevisiae* (Hutter and Oliver, 1998).

2.3. Yeast Nutrition

Yeasts acquire essential elements from their growth environment from simple food sources. Elements C, H, O, N, P, K, Mg and S are needed to be available at the macronutrient level (approx. 10^{-3} M) and trace elements are needed at the micronutrient level (approx. 10^{-6} M) (Walker, 1998).

S. cerevisiae is a chemoorganotrophic organism obtaining carbon and energy from compounds in fixed, organic linkage. It can utilize the sugar types glucose, fructose, mannose, galactose, sucrose and maltose by fermentative or respirative metabolism. Other carbon sources (e.g. ethanol, acetate) can by utilized as respiratory substrates. Yeasts cannot fix molecular nitrogen, thus simple inorganic nitrogen sources such as ammonium salts are widely utilized. Ammonium sulphate is a commonly used nitrogen source in yeast growth media since it also provides a source of assimilable sulphur. Yeast require sulphur principally for the biosythesis of sulphur-containing amino acids (Walker, 1998).

Oxygen is an important yeast growth factor and *S. cerevisiae* is unable to grow well in the complete absence of oxygen since oxygen is required for biosythesis of growth-related molecules like sterols and unsaturated fatty acids besides being a substrate for respiratory enzymes. Elemental hydrogen is available from carbohydrates and other sources. Hydrogen ions (protons) are very important since variations in both extracellular and intracellular pH affect the growth and metabolism of yeast cells. In growing cells, intracellular pH is regulated to within a narrow range around 5.25 through the action of the plasma membrane proton-pumping ATPase because pH level regulates yeast nutrient uptake and subsequent metabolism. However, extracellular pH variations have limited effect on the yeast cytosolic pH unless the cell is exposed to extreme values. Phosphorus is present in nucleic acids and in phospholipids and therefore it is essential for *S. cerevisiae*. A significant contribution to the negative charge of the yeast cytoplasm is due to presence of inorganic phosphates and the phosphate groups in organic compounds (Walker, 1998).

Yeast's requirements for minerals are similar to those of other cells, with a supply of potassium, magnesium and several trace elements being necessary for growth. Growth factors are organic compounds required in very low concentrations for specific catalytic or structural roles in yeast, but are not utilized as energy sources. Yeast growth factors include: Vitamins, purines and pyrimidines, nucleosides and nucleotides, amino acids, fatty acids, sterols and other miscellanous compounds (Walker, 1998).

Yeasts absorb the nutrients mentioned above through the cell membrane as low molecular weight compounds dissolved in water. Entry of water, which is essential for growth, is controlled at the level of the cell envelope. Water permeability is rate-limited by passage through both the yeast cell wall and plasma membrane. The driving force for water inflow into fungal cells is the turgor potential, being the difference between the cells osmotic potential and its water potential. The osmotic potential in *S. cerevisiae* is created by the accumulation of solutes from the growth environment. Under steady-state conditions, yeast may be considered as growing at relatively constant water potential (Walker, 1998).

2.4. Central Carbon Metabolism of *S. cerevisiae*

Most yeasts utilize sugars as their preferred carbon and energy sources. The sequence of enzyme-catalysed reactions that oxidatively convert glucose to pyruvic acid in the yeast cytoplasm is known as glycolysis which is given in Figure 2.1 (the enzymes. Glycolysis provides yeast with energy, together with precursor molecules and reducing power for biosynthetic pathways. The key regulatory enzymes in glycolysis are irreversible phosphofructokinases (*PFK*) and pyruvate kinase (*PYC*) whose activity is influenced by numerous effectors, including ATP (Walker, 1998).

With regard to fermentation, several biotechnologically important yeasts including *S. cerevisiae* are fermentative and, as such, are defined as organisms which use organic substrates anaerobically as electron donor, electron acceptor and as carbon source. During alcoholic fermentation of sugars, yeast re-oxidizes NADH to NAD in terminal step reactions from pyruvate. In the first of these reactions pyruvate is decarboxylated to acetaldehyde, catalysed by pyruvate decarboxylase (*PDC*), and then to reduced to ethanol

catalysed by alcohol dehydrogenase (*ADH*). Regeneration of NAD is necessary to maintain the redox balance and prevent the stalling of glycolysis. An alternative means of replenishing NAD is the pathway from dihydroxy acetone phosphate (DAHP) to glycerol.

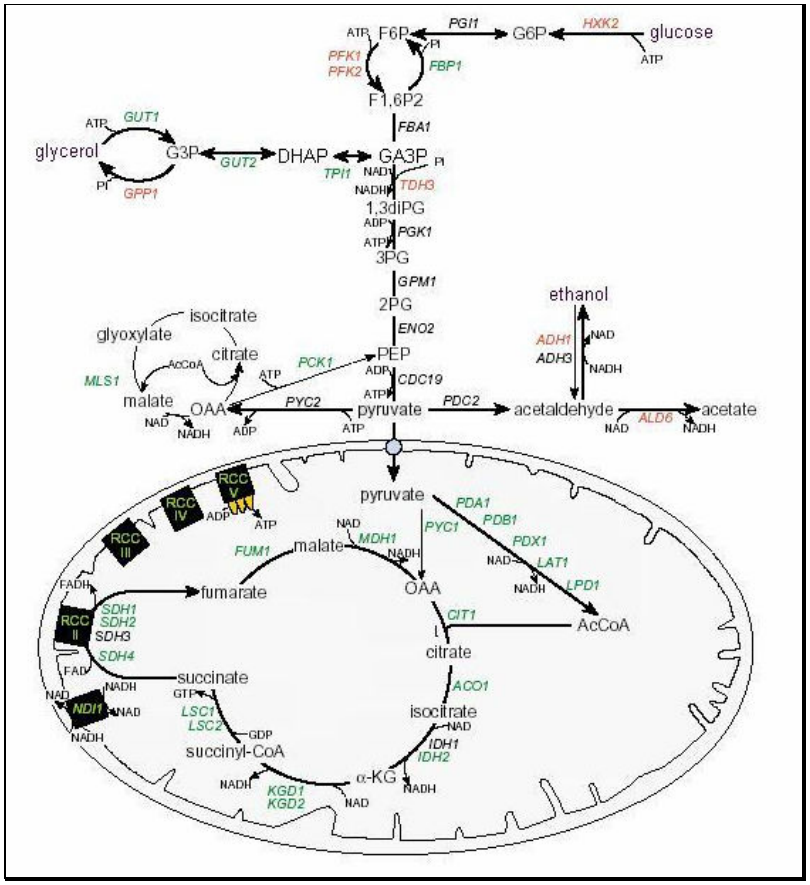

Figure 2.1. Central carbon metabolism of *S. cerevisiae* (Steinmetz *et al.*, 2002).

Glycerol production maintains the redox balance of the cell but does so at the expense of carbohydrate conversion to ethanol. Succinate, together with glycerol, represents the major secondary fermentation product in *S. cerevisiae*. Succinate is thought to be synthesized and secreted by yeasts, either following limited operation of the citric acid cycle or by reductive linear pathways involving some citric acid cycle enzymes (Walker, 1998).

In respiratory metabolism, glycolysis is followed by the citric acid cycle, the electron transport chain and oxidative phosphorylation, which collectively extract most of the energy (in form of ATP) from glucose. Also other carbon sources can be respired rather

than fermented in addition to glucose. Those substrates which are respired by yeast cells include: pentoses, sugar alcohols, organic acids, aliphatic alcohols, hydrocarbons and aromatic compounds.

Considering glucose respiration, in the presence of oxygen and absence of repression, pyruvate enters the mitochondrial matrix where it is oxidatively decarboxylated to acetyl CoA. This reaction is catalysed by the pyruvate dehydrogenase multi enzyme complex (*PDH*) which acts as the link between glycolysis and the citric acid cycle. The activated acetyl unit is then completely oxidized to two molecules of CO_2 by the TCA cycle (tricarboxylic acid, citric acid or Krebs cycle), which is the final common pathway for the oxidation of sugars and other carbon sources in yeast.

During the TCA cycle, hydrogens are transferred, in processes mediated by dehydrogenase enzymes, to the redox carriers NAD and FAD, which become reduced. The reduced carriers are reoxidized and oxygen is reduced to water via the electron transport chain on the inner mitochondrial membrane. The energy released by the transfer of electrons is used to synthesize ATP by oxidative phosphorylation. Glycolysis and TCA cycle in yeasts therefore encompasses oxidation reactions which produce a small amount of ATP by substrate-level phosphorylation and NADH. The latter is then processed by electron transport to oxygen, as the terminal electron acceptor, to yield much more ATP.

The electron transport chain pumps protons out of the mitochondrial matrix across the inner mitochondrial membrane to create a transmembrane proton gradient and a membrane potential difference, which together comprise the “proton-motive force”. This is the driving force used to synthesize ATP (by H^+- transporting ATPase). The total yield of ATP for complete oxidation of glucose by yeast cells is around 30. This is because each pair of electrons in NADH generates about 1.5 ATP while succinate oxidation yields about 15 ATP. The proportion of energy which is not retained for use by the yeast cell as ATP is largely dissipated as metabolic heat.

NADH produced in glycolysis cannot be oxidized directly within yeast mitochondria due to impermeability of these organelles to NADH. Cytoplasmic shuttle processes enable reduced cofactors to enter mitochondria. For example, the glycerophosphate shuttle utilizes

NADH to reduce dihydroxyacetone to glycerol 3-phosphate and malate shuttle utilizes NADH to reduce oxaloacetate to malate. These molecules then enter the mitochondrion where enzymes oxidize them to yield reduced cofactors which in turn are oxidized by electron transport chain (Walker, 1998).

2.5. Regulation of Sugar Metabolism

Biochemical pathways in yeasts may be regulated at various levels. These include: enzyme synthesis (e.g. induction, repression and derepression of gene expression), enzyme activity (e.g. allosteric activation, inhibition or interconversion of isoenzymes) and cellular compartmentalization (e.g. mitochondrial localization of respiratory enzymes). The following discussion will focus primarily on external factors which influence respiratory and fermentative metabolism in *S. cerevisiae*.

The Pasteur Effect relates oxygen with the kinetics of yeast sugar metabolism and states that under anaerobic conditions, glycolysis proceeds faster than it does under aerobic conditions. Alternatively, it may be defined as a suppression of fermentation by oxygen. However, this phenomenon is only observable when glucose concentrations are low (e.g. below around 5 mM in *S. cerevisiae*) or under certain nutrient-limited conditions. It is associated with a decreased affinity for sugar uptake under aerobic conditions (Walker, 1998).

If the concentration of available glucose is high, the Pasteur Effect in *S. cerevisiae* is no longer operable. Then the Crabtree effect comes into consideration. This phenomenon (also referred to as the glucose effect or contra-effect Pasteur) relates glucose concentration with the particular catabolic route adopted by glucose-sensitive yeasts (like *S. cerevisiae*) in the presence of oxygen and states that, even under aerobic conditions, fermentation predominates over respiration. Thus, even though oxygen may be present, NADH generated during glycolysis is mainly oxidized by fermentation, rather than by respiration.

In *S. cerevisiae*, glucose repression is thought to be due to glucose repressing respiratory enzyme synthesis and/or inactivating respiratory enzymes and sugar transport activity. In addition to glucose-repression mechanisms, normal yeast mitochondrial

structures are disrupted when glucose levels are high. For example, inner membranes and cristae disappear at high glucose concentration, and then re-appear once aerobic metabolism replaces alcoholic fermentation. Alteration of cellular compartmentalization in this manner clearly has profound metabolic consequences in yeast cells (Walker, 1998).

The short-term Crabtree effect is observed as a sudden fermentative response when excess sugar is added to a non-fermenting yeast culture, whereas a long-term Crabtree effect is manifest by aerobic fermentation under fully adapted physiological steady-state conditions. The latter has been explained on the basis of the limited respiratory capacity of Crabtree-positive yeasts; and the former on the basis of saturation of respiration leading to overflow of pyruvate. The Crabtree effect is not restricted to growth on glucose since *S. cerevisiae* will also aerobically ferment fructose. In the presence of mannose or galactose, aerobically grown *S. cerevisiae* will simultaneously respire and ferment these sugars (Walker, 1998).

Glucose repression in yeasts like *S. cerevisiae* describes a long-term regulatory adaptation to degrade glucose exclusively to ethanol and CO_2. Therefore, when *S. cerevisiae* is grown aerobically on high concentrations of glucose, fermentation will account for the bulk of glucose consumption. In batch culture, however, when the levels of consumed glucose decline, cells will gradually become derepressed, resulting in induction of respiratory enzyme synthesis. This, in turn, results in oxidative consumption of accumulated ethanol when cells enter a second phase of growth known as diauxie (Walker, 1998).

The Crabtree effect may also be due to a saturation of the limited respiration capacity of yeast cells. Thus, glucose-sensitive (Crabtree-positive) yeasts, like *S. cerevisiae*, may possess a limited oxidative capacity when grown on glucose which leads to an overflow reaction at pyruvate. When the respiratory capacity is saturated, ethanol is formed (Walker, 1998).

The regulation of respiration and fermentation is fundamental to the success of several industrial processes, that exploit yeast metabolism. In *S. cerevisiae*, optimization of respiration is important for the production of yeast biomass (e.g. for the food industry),

while optimization of fermentation is important for potable and industrial ethanol production (Walker, 1998).

2.6. Oxidative Phosphorylation

The citric acid cycle accounts for most of the total oxidation of carbon compounds in yeast cells, and its end products are CO_2 and high-energy electrons, which pass via NADH and $FADH_2$ to the respiratory chain. None of the reactions leading to NADH or $FADH_2$ production makes direct use of molecular oxygen; only in the final catabolic reactions that take place on the mitochondrial inner membrane, directly oxygen is consumed. Nearly all of the energy available from catabolism of sugar substrates in the earlier stages of oxidation is saved in the form of high-energy electrons removed from substrates by NAD^+ and FAD. These electrons, carried by NADH and $FADH_2$ are then combined with molecular oxygen by means of respiratory chain (Figs. 2.2). Because the large amount of energy released is harnessed by the enzymes in the inner membrane to derive the conversion of ADP + P_i to ATP, the term oxidative phosphorylation is used to describe this last series of reactions (Alberts *et al.*, 1994; Trumpower and Gennis, 1994).

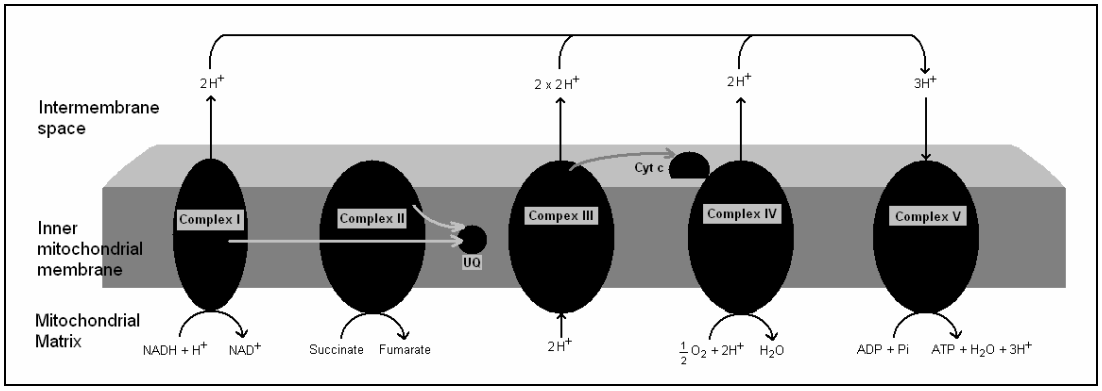

Figure 2.2. Respiratory chain (adapted from Mandavilli *et al.*, 2002).

2.7. The Respiratory Chain

There are several types of electron carriers and, five large membrane-bound enzyme complexes, in the respiratory chain, which are embedded in the inner mitochondrial

membrane. The cytochromes, iron-sulphur proteins, ubiquinone and flavins are the major electron carriers in the respiratory chain.

The protein complexes in respiratory chain of *S. cerevisiae* cells are depicted in Figure 2.2. These complexes are embedded in the inner mitochondrial membrane and composed of 44 proteins (Table 2.2), 7 of which are encoded by mitochondrial genes, and the rest of them are encoded by nuclear genes. Each of these complexes was studied in detail, their structure and to an extent their regulation are well known (Güldener *et al.*, 2005).

Table 2.3 summarizes all the genes that encode subunits of complexes of respiratory chain in *S. cerevisiae*. In this list, ORFs (open reading frames) of these genes are also given, the ones with initial 'Q' are the ORFs on mitochondrial DNA, the rest with initial 'Y' are nuclear ORFs.

2.7.1. NADH Dehydrogenase: Complex I

The first complex, NADH-dehydrogenase complex does not exist in *S. cerevisiae*, it is replaced by the enzyme NADH-ubiquinone-6 oxido-reductase (Ndi1p, 57 kDa). This enzyme catalyses the oxidation of NADH to NAD^+, ubiquinone is reduced to ubiquinol in the same reaction (Güldener *et al.*, 2005).

2.7.2. Succinate Dehydrogenase: Complex II

The second complex, succinate dehydrogenase (fumarate reductase) is composed of four proteins in *S. cerevisiae*, which are: a membrane-anchoring protein (encoded by *SDH4*, precursor weight of whose product is 20 kDa), a flavoprotein (Sdh1p, 70 kDa), an iron-sulfur protein (Sdh2p, 27 kDa) and a cytochrome b (Sdh3p, 22 kDa). This complex is responsible for transfering electrons from succinate to ubiquinone. That is, succinate is converted to fumarate and ubiquinone is reduced to ubiquinol. Succinate dehydrogenase is the only TCA cycle enzyme to be bound to the inner mitochondrial membrane (Güldener *et al.*, 2005).

Table 2.2. Substrates and products of reactions taking place in respiration chain complexes (Kegg, 2005).

		Substrates		Products	
Complex	Enzyme	Electron donor	Electron acceptor	Oxidized	Reduced
I	NADH dehydrogenase (ubiquinone)	NADH	Ubiquinone	NAD+	Ubiquinol
II	Succinate dehydrogenase (ubiquinone)	Succinate	Ubiquinone	Fumarate	Ubiquinol
III	Ubiquinol--cytochrome c reductase	Ubiquinol	Ferri-cytochrome c	Ubiquinone	Ferro-cytochrome c
IV	Cytochrome c oxidase	Ferro-cytochrome c	O_2	Ferri-cytochrome c	H_2O

2.7.3. Cytochrome bc1: Complex III

The third complex, cytochrome bc1 (or ubiquinol-cytochrome c reductase) complex is composed of 10 subunits. It accepts electrons from the Rieske iron-sulfur protein and transfers electrons to cytochrome c.

Only one of the components of Complex III is encoded by mitochondrial DNA: Cobp or Ctybp (44kDa). The other subunits are all encoded by nuclear genes and are translated on cytoplasmic ribosomes as precursors that are proteolytically processed in one or more steps during transport into their destinated internal compartment of the mitochondria. The *COB* gene of yeast contains at least two and in some cases as many as five introns. Some introns code for splicing factors that promote the excision of their own coding sequences (Tzagoloff, 1995). Cbp6p and Cbp7p are translational activators of cytochrome b, Cbp6p also takes part in pre-mRNA processing. Cbp2p and Cbt1p are known to function in mRNA splicing. Cbp1p performs 5' end processing as well as stabilizing *COB* mRNA. Mrs1p, Suv3p and Nam1p regulates both *COX1* (Complex 4 component) and *COB* transcripts.

Table 2.3. Proteins involved in respiration chain complexes

ORF	Gene	Function
NADH-dehydrogenase complex (ubiquinone, complex I)		
YML120c	*NDI1*	NADH-ubiquinone-6 oxido-reductase
Succinate dehydrogenase complex (complex II)		
YDR178w	*SDH4*	succinate dehydrogenase membrane anchor subunit for SDH2P
YKL141w	*SDH3*	cytochrome b560 subunit of respiratory complex II
YKL148c	*SDH1*	succinate dehydrogenase flavoprotein precursor
YLL041c	*SDH2*	succinate dehydrogenase iron-sulfur protein subunit
Cytochrome bc1 complex (Ubiquinol-cytochrome c reductase complex, complex III)		
Q0105	*COB*	ubiquinol-cytochrome-c reductase subunit (cytochrome B)
YBL045c	*COR1*	ubiquinol--cytochrome-c reductase 44K core protein
YDR529c	*QCR7*	ubiquinol--cytochrome-c reductase subunit 7
YEL024w	*RIP1*	ubiquinol--cytochrome-c reductase iron-sulfur protein precursor
YFR033c	*QCR6*	ubiquinol--cytochrome-c reductase 17K protein
YGR183c	*QCR9*	ubiquinol--cytochrome-c reductase subunit 9
YHR001w-a	*QCR10*	ubiquinol--cytochrome-c reductase 8.5 kDa subunit
YJL166w	*QCR8*	ubiquinol--cytochrome-c reductase chain VIII
YOR065w	*CYT1*	cytochrome-c1
YPR191w	*QCR2*	ubiquinol--cytochrome-c reductase 40KD chain II
Cytochrome c oxidase (complex IV)		
Q0045	*COX1*	cytochrome-c oxidase subunit I
Q0250	*COX2*	cytochrome-c oxidase subunit II
Q0275	*COX3*	cytochrome-c oxidase chain III
YDL067c	*COX9*	cytochrome-c oxidase chain VIIA
YGL187c	*COX4*	cytochrome-c oxidase chain IV
YGL191w	*COX13*	cytochrome-c oxidase chain VIa
YHR051w	*COX6*	cytochrome-c oxidase subunit VI
YLR038c	*COX12*	cytochrome-c oxidase, subunit VIB
YLR395c	*COX8*	cytochrome-c oxidase chain VIII
YMR256c	*COX7*	cytochrome-c oxidase, subunit VII
YNL052w	*COX5A*	cytochrome-c oxidase chain VA precursor
YIL111w	*COX5B*	cytochrome-c oxidase chain VB precursor
F0/F1 ATP synthase (complex V)		
Q0080	*AAP1*	F1F0-ATPase complex, F0 subunit 8
Q0085	*ATP6*	F1F0-ATPase complex, FO A subunit
Q0130	*OLI1*	F1F0-ATPase complex, F0 subunit 9
YBL099w	*ATP1*	F1F0-ATPase complex, F1 alpha subunit
YBR039w	*ATP3*	F1F0-ATPase complex, F1 gamma subunit
YDL004w	*ATP16*	F1F0-ATPase complex, F1 delta subunit
YDL181w	*INH1*	inhibitor of mitochondrial ATPase
YDR298c	*ATP5*	F1F0-ATPase complex, OSCP subunit
YDR322c-a	*TIM11*	subunit E of the dimeric form of mitochondrial F1F0-ATPase
YDR377w	*ATP17*	ATP synthase complex, subunit f
YJR121w	*ATP2*	F1F0-ATPase complex, F1 beta subunit
YKL016c	*ATP7*	F1F0-ATPase complex, FO D subunit
YLR295c	*ATP14*	F1F0-ATPase complex, subunit h
YML081c-a	*ATP18*	gene for subunit i/j of the mitochondrial F1F0-ATP synthase
YOL077w-a	*ATP19*	subunit K of the dimeric form of mitochondrial F1F0-ATP synthase
YPL078c	*ATP4*	F1F0-ATPase complex, F0 subunit B
YPL271w	*ATP15*	F1F0-ATPase complex, F1 epsilon subunit
YPR020w	*ATP20*	subunit G of the dimeric form of mitochondrial F1F0-ATP Synthase

Suv3p stabilizes both of these mRNAs. Mrs1p functions in the excision of intron bI3 in *COB* mRNA and intron aI5b in *COX1* mRNA. Cobp has protein-protein interaction with Qcr9p and suppresses Core2 (Güldener *et al.*, 2005).

The protein Core1 (expressed from gene *QCR1* or *COR1*, precursor size is 50 kDa), an essential subunit of bc1 complex, is required to convert apocytochrome b to mature cytochrome b. It may also mediate formation of the complex between cytochromes c and c1. It is regulated by heme although it is not a heme protein. The protein Core 2 (Qcr2p or Cor2p, 40.5 kDa) is also regulated by heme, although it is not a heme protein. It is a component of the bc1 complex and is also required for assembly (Güldener *et al.*, 2005).The protein encoded by *QCR6* or *COR3* (17 kDa), is another component of bc1 complex. It may help bind Cyt1p to Cyc1p (cytochromes c and c1). It has a protein-protein interaction with Sin4p. The product of the *QCR7* or *COR4* (15 kDa) gene plays a role in formation of complex bc1 and has a protein-protein interaction with Duo1p. It binds to ubiquinone and stabilizes ubisemiquinone radicals. It forms an initial core complex with Qcr8p that is essential for subsequent assembly of mature complex. Transcription of *QCR8* or *COR5* (which encodes a protein of size 11 kDa) needs the HAP2/3/4 complex for rapid induction during transition from repressed to derepressed conditions. Abf1p may act in coordination with the HAP2/3/4 complex while Cpf1p is a negative regulator modulating the induction response. *QCR8* has an upstream binding site for Mig1p (Güldener *et al.*, 2005).

Qcr9p (7.5 kDa) is also essential for formation of a fully functional bc1 complex. It interacts with RIP1p and Cobp. It contains an intron that is nearly identical to the intron of *COX4* suggesting coordinated regulation of splicing. Qcr10p (8.6 kDa) is a subunit of the bc1 complex, whose presence is probably required for stable association of the Rieske iron-sulfur protein (Rip1p). The Rieske iron-sulfur protein (Rip1p, 23.4 kDa) is another component of the bc1 complex. This protein is known to interacti with Gle1p and Gle2p. The last component is cytochrome c1 (Cyt1p or Ctc1p, 34 kDa). Its function is to accept electrons from the Rieske protein and transfer them to cytochrome c. It contains a heme, and has an upstream binding site for Mig1p (Güldener *et al.*, 2005).

The catalytic subunits of the bc1 complex are Cobp, Cyt1p and Rip1p. These proteins, together with the other non-catalytic subunits (Core1, Core2, Qcr6p, Qcr7p, Qcr8p, Qcr9p, Qcr10p), assemble to form an enzymatically active complex. The analysis of the steady-state levels of these subunits has suggested that the assembly pathway of this complex occurs in a coordinated fashion, involving the formation of specific assembly intermediates. According to this model, cytochrome b initially forms a subcomplex with Qcr7p and Qcr8p, which subsequently joins with the Core1 and Core2 proteins. Cytochrome c1, on the otherhand, is proposed to form another subcomplex with Qcr6p and Qcr9p. Formation of each of these subcomplexes ensures stability against proteolytic attack for the individual subunits contained within them. The cytochrome b and cytochrome c1 subcomplexes subsequently unite to form a 'cytochrome bc1 precomplex' prior to the assembly of the Rieske FeS protein and, presumably, the non-essential subunit Qcr10p. (Stuart *et al.*, 1999).

2.7.4. Cytochrome c Oxidase: Complex IV

The fourth complex, cytochrome c oxidase is composed of 11 subunits and this complex is responsible for transferring electrons from cytochrome c to molecular oxygen. The reactions taking place are summarized in Table 2.3. Cytochrome c oxidase may exist either in a monomer or a dimer in the mitochondrial membrane. There is also evidence that it forms a dimer with Complex III (Stuart *et al.*, 2000, Berden *et al.*, 1998). Stability and splicing of mRNA of subunit 1 (which is encoded by the mitochondrial gene *COX1* or *OXI3*), depends on Pet309p, Nam1p, Suv3p. The gene of subunit II (*COX2* or *OXI1*) is also mitochondrially encoded, and transcriptionally and post-transcriptionally regulated by Nam1p and Mss2-1p, respectively. Also, Pet111p promotes its translation. The third subunit (mitochondrially encoded Cox3p or Oxi2p) is known to be transcriptionally regulated by Nam1p. The expressions of the fourth (*COX4*) and sixth subunits (*COX6*) are known to be induced after shift to aerobic conditions and 1μM oxygen is a threshold for expression. Proteins encoded by *COX1*, *COX2*, *COX3*, *COX4* accumulate in an interdependent way, whereas the accumulation of Cox6p appears to be independent from the other four (Güldener *et al.*, 2005).

The fifth subunit of the fourth complex in the respiratory chain has two interchangeable isoforms in *S. cerevisiae* (Cox5Ap and Cox5Bp). These isoforms differentially affect the catalytic properties of the holoenzyme expressed under aerobic conditions, and *COX5A*/*COX5B* are co-regulated with *CYC1*/*CYC7* which are the isoforms in third complex, expressed under aerobic/anaerobic conditions, respectively (Poyton *et al.*, 1997).

Cox12p (10 kDa) and Cox13p (15 kDa) are known as subunits VIa and VIb of complex four, respectively. Cox13p is not essential for assembly of cytochrome c oxidase and it is not a structural part, but Cox12p is essential for assembly though it is not necessarily part of the assembled enzyme complex. Cox9p (7 kDa), is known as subunit VIIa, and plays a role in holoenzyme assembly or stability. Cox7p (7 kDa), is known as subunit VII and is involved in assembly of the complex. Both Cox9p and Cox7p are induced after shift to aerobic conditions and 1 μM oxygen is a threshold for expression (Güldener *et al.*, 2005).

2.7.5. ATP Synthase: Complex V

ATP synthase, which catalyses the conversion of ADP + P_i to ATP is the fifth complex. In *S. cerevisiae*, mitochondrial ATPase is of F type. This F_0/F_1 ATP synthase complex is composed of 18 types of proteins, three of which are encoded by mitochondrial DNA. Atp8p or Aap1p (6 kDa), Atp6p or Oli2p or Oli4p (29kDa), ATP9p or Oli1p (8 kDa) are encoded mitochondrially and they are subunits of F_0. Levels of the mRNAs of *ATP9* and *ATP6* are reduced in null mutants of *nam1*, i.e. they are transcriptionally regulated by Nam1p. *ATP6* mRNA stability also depends on Pet309p (Güldener *et al.*, 2005).

F-type ATPases have two components, CF(1)- the catalytic core and CF(0)- the membrane proton channel. F_0/F_1 ATP synthase may exist as a dimer or a monomer; dimerization appears to confer a stability advantage on the complex. The dimer contains three additional dimer-specific proteins Tim11p or Atp21p (11 kDa), Atp19p (7.5 kDa), Atp20p (13 kDa) (subunits E, K and G of the dimeric form, respectively) which are not required for the catalytic activity of ATPase. Tim11p and Atp20p play an essential role in

the formation of dimers of the F_0/F_1 ATP synthase complex. Presence of Atp20p enhances the stability of Tim11p, and Tim11p is essential for the stable expression of *ATP20* and *ATP19* (Güldener *et al.*, 2005).

Atp1p (59 kDa), Atp2p (55 kDa), Atp3p (34 kDa) and Atp16p (17 kDa), Atp15p (7 kDa) are the alpha, beta, gamma, delta and epsilon subunits of the F_1 complex. Atp1p is a regulatory subunit while Atp2p is an intramolecular chaperone and a catalytic unit. ATP3p regulates ATPase activity and the flow of electrons through F_0 complex, and play role in the assembly/stability of F_1 (Güldener *et al.*, 2005).

Atp5p (23 kDa), is the OSCP (oligomycin-sensitivity-conferring-protein) subunit of F_1 complex and is required for assembly of the catalytic F_1 sector. It either transmits conformational changes from F_0 into F_1 or is implicated in proton conduction. Atp17p (11 kDa) is the subunit f, mitochondria of its disruption mutant contain active F_1 but the subunits f, 6 and 8 are absent. Atp4p (20 kDa) and Atp7p (20 kDa) are the b and d subunit of F_0, respectively. In null mutants *atp7*, F_0 lacks both subunits 4 and 6, cytochrome oxidase activity is also low (Güldener *et al.*, 2005).

Atp14p (14 kDa) is the subunit h, its disruptants lack subunit h and subunit 6 and show no ATPase activity. Atp18p (7 kDa), which is the subunit i/j, is not a structural part of the complex but, rather is involved in oxidative phosphorylation. Inh1p (10 kDa) is an inhibitor of mitochondrial ATPase. It forms a one-to-one complex with ATPase to inhibit enzyme activity completely (Güldener *et al.*, 2005).

2.8. Literature Search on the Genes Investigated in This Study

The deletion mutants of six genes are investigated in this study. The functions of the well-known regulators Hap4p and Mig1p were studied previously in detail using both direct and high-throughput methods. The *BCS1*, *OXA1* and *MBA1* genes were reported to have regulatory functions but limited information on the effects of their deletions was known. *RIP1* encodes Rieske iron-sulfur component of Complex III in the respiratory chain. It is not observed to have any regulatory function. The following sections summarize the information in the literature about these genes and their products.

2.8.1. *HAP4*

The function of the Hap4p protein was determined to be the regulation of the respiration related-genes (DeRisi *et al.,* 1997). The activation mechanism of HAP2/3/4/5 protein complex was reviewed by Gancedo (1998). The physiology of cells with *HAP4* overexpression (Blom *et al.*, 2004) and the transcriptome profile of *hap4* deletion mutants (Buschlen *et al.*, 2003) were also investigated. *Hap4* deletion mutants were reported to be respiratory deficient (Steinmetz *et al*., 2001) and deletion of *HAP4* caused downregulation of the respiration-related genes. On the contrary, these genes were expressed at higher levels in *HAP4*-overexpressing strains growing under aerobic conditions. In addition, increased levels of respiratory capacity were observed due to overexpression of *HAP4* (Blom *et al.*, 2004). The analysis of the pathways that are activated and inactivated in *hap4* deletants were investigated under glucose- and ammonium-limited conditions in this study. The genes that mediate the changes in *hap4* phenotype (biomass production, glucose consumption and ethanol production) and the pathways involving these genes were also reported in this study. Results of the present experiments generally agree with the information in the literature, but include some novel findings.

2.8.2. *MIG1*

Mig1p is involved in carbon catabolite repression of expression of genes encoding the enzymes of central carbon metabolism in the yeast *S. cerevisiae*. The proposed mechanism of *MIG1* control on repressed genes depends on the glucose level. Mig1p is dephosphorylated by protein phosphatases Reg1p, Reg2p and Glc1p when glucose is present and repression of target genes takes place. When glucose is absent, phosphorylation of Mig1p by protein kinase Snf1p results in derepression of target genes (de Vit *et al*., 1997). *MIG1* was reported to be auto-regulated. Mechanisms of *MIG1*-mediated glucose regulation were reviewed and effects of *mig1Δ* deletion on peripheral and central metabolisms were discussed by Klein *et al.* (1998). No effect of the *mig1Δ* deletion on the fluxes of central metabolic reactions was observed when the cells were grown in chemostats at low-glucose concentration (Gombert *et al.*, 2001), which may indicate that *MIG1* does not affect metabolism at low concentration of glucose because glucose repression does not take place. Also, fluxes were reported to be changing

insignificantly in a *mig1Δ* deletant at even higher glucose concentrations (Raghevendran *et al*., 2004). As it will be illustrated in the following sections, the results of the present study show that deletion of *MIG1* has effects on the expression of many genes in cells of chemostat cultures operated under glucose-limited conditions.

2.8.3. *OXA1*

Oxa1p is a component of the conserved Oxa1/YidC/Alb3 protein family which acts as a membrane insertion machinery, and is involved as a general protein insertion site in the mitochondrial inner membrane using proteins encoded by both the nuclear and mitochondrial genomes of *S. cerevisiae* (Stuart, 2002). Oxa1p is required for mitochondrially encoded Cox2p to be exported to the mitochondrial intermembrane space, but its function is not limited to transport of Cox2p (Hell *et al.*, 1998). Absence of Oxa1p was reported to increase levels of Pdr3p and Pdr5p, while assembly of Complex IV and F_0 sector of ATPase do not take place. Pdr3p and Pdr5p are proteins participating in retrograde signalling from the mitochondria to the nucleus (Lemaire *et al.*, 2000). *oxa1Δ* deletion mutants was reported to be deficient in assembly of Complex IV and Complex V and deficient in respiration (Altmura *et al.*, 1996). However, Oxa1p was not found to be essential for transport of ATPases (Lemaire *et al.*, 2000). Enzyme activity measurements of Complex III, Complex IV and Complex V showed that *oxa1Δ* deletion mutants have only 19%, 0% and 16%, respectively of the activities in wildtype cells (Lemaire *et al.*, 2004). These three complexes have both mitochondrially and nuclearly encoded components. My results verify that deletion strain *oxa1Δ/oxa1Δ* suffers respiratory deficiency and indicate that Oxa1p may be an activator of mitochondrially encoded components of the respiratory chain.

2.8.4. *MBA1*

Mba1p is thought to have a similar function as Oxa1p. Mba1p can act either in coorporation with Oxa1p or may accomplish a similar function independently in protein insertion machinery of *S. cerevisiae* mitochondria (Preuss *et al.*, 2001). Strains with the *mba1Δ* deletion suffer partial respiratory deficiency, thus its protein may be involved in assembly of respiratory chain (Rep and Grivell., 1996). My results verify that homozygous

mba1Δ deletants suffer partial respiratory deficiency. However, the expression behaviour of *mba1Δ* deletion mutants is similar to *mig1* deletion mutants. Moreover, *MBA1* deletion causes *MIG1* expression to be repressed. Thus, it may be concluded that Mba1p is an activator of *MIG1* and deletion of these two genes cause the same expression profiles.

2.8.5. *RIP1* and *BCS1*

Rip1p is the Rieske iron-sulfur subunit of Complex III and *BCS1* is defined as a gene necessary for a functional Complex III in the respiratory chain of *S. cerevisiae* (Nobrega *et al.*, 1992). Deletion of *bcs1* results in reduced levels of Rip1p. It was stated that *BCS1* regulation of *RIP1* is not likely to be transcriptional or translational. Instead, Bcs1p may be involved in post-translational modification of the apoprotein. Phenotypes of *rip1Δ/rip1Δ* and *bcs1Δ/bcs1Δ* deletion mutants are similar, mainly other components of the complex are not affected by these deletions (Nobrega *et al.*, 1992 and Crivellone *et al.*, 1988).

2.8.6. Respiratory-Deficient Strains

In respiratory-deficient yeast cells, glucose cannot be metabolised through oxidative phosphorylation, thus fermentation to ethanol is enhanced. The TCA cycle is not completed, succinate can not be oxidized to fumarate. In order to compensate for production of glutamate from α-ketoglutarate, the pathways producing oxaloacetate (OAA) and acetyl-CoA are induced. Also the first three steps of the TCA cycle and glyoxylate cycle activity is increased (Butow and Avardhani, 2004). My results verify that the first three steps of the TCA cycle are induced under respiratory-deficient conditions.

The deletion mutations causing respiratory deficiency in the present study are as follows: Homozygous deletion mutants *hap4Δ/hap4Δ*, *bcs1Δ/bcs1Δ*, *oxa1Δ/oxa1Δ*, and *rip1Δ/rip1Δ*. Also homozygous deletion mutant *mba1Δ*/*mba1Δ* and heterozygous deletion mutants *HAP4*/*hap4Δ* and *RIP1*/*rip1Δ* suffer partial deficiency of respiration.

2.9. Functional Genomics and Systems Biology

Functional genomics is the branch of biotechnological research that aims the determination of functions of ORFs that are not yet identified. The functions of many genes revealed in large-scale sequencing projects can be inferred through nucleotide similarity with the sequences of genes of known function determined through traditional empirical methods. However, there still remains a large number of predicted open reading frames (ORFs) that have no assigned function based on sequence similarity. Thus, empirical methods of function determination are required. Functional elucidation of genes can be pursued through the systematic perturbation of gene expression followed by quantitative and qualitative analyses of gene expression products including mRNA, protein, and metabolite levels.

Systems biology is similar to functional genomics in its approach, but is slightly different in its objectives. Systems biology aims to simultaneously monitor all biological processes operating as an integrated system. Through the study of systems, one can begin to visualize how individual pathways or metabolic networks are interconnected. This approach is based on solid theoretical frameworks and uses computer modeling to explain experimental observations. It is envisioned that systems biology will be of great value for directing metabolic engineering strategies because modifications to the expression of single genes do not always bring about the predicted or desired effects due to cross-talk between pathways (Sumner *et al.*, 2003).

After the completion of the genomic sequencing of organisms, integrative post-genomic studies and the systems biology approach have emerged with the aim of developing a more complete understanding of cell physiology. Attempts at data integration for the model organism, *Saccharomyces cerevisiae* were reviewed recently (Castrillo and Oliver, 2004). Experimental designs that involve (a) perturbations to elucidate the response of the cell under various conditions, (b) collection of high-throughput data at different functional genomic levels and (c) the use of bioinformatics for integrating data from all three levels of analysis (transcriptome, proteome, and metabolome) constitute the three major steps common to all integrative studies.

It is possible to design systems biology experiments in a hypothesis-driven manner, such that the designed perturbations provide the information of interest. Alternatively, question-driven discoveries may be made by observing the effects of an intuitively chosen modification and making use of the extracted information to generate new ideas and hypotheses (Lockhart and Winzeler, 2000).

Transcriptome (mRNA synthesis) data from *S. cerevisiae* growing in chemostats on a glucose medium under either carbon, nitrogen, phosphorus or sulphur limitation allowed detection of the genes that were affected by the different nutrient limitations (Boer *et al.,* 2003). The genes that were co-regulated under glucose, ethanol, ammonium or phosphate limitation were identified, and genes from the same pathway were shown to be clustered together (Wu *et al.*, 2004). Responses to modifications in the growth medium and/or the dilution rate allowed the identification of genes that enable the cells to adapt to various growth conditions (Hayes *et al.*, 2002).

Perturbations can also be introduced by genetic, rather than physiological, means, e.g. by gene deletion. Yeast cells carrying gene deletions have been investigated for various purposes: (a) functional analysis based on discrimination of mutants via metabolic fingerprints (Raamsdonk *et al.,* 2001) or footprints (Allen *et al.*, 2003), (b) selection of genes encoding organelle-specific proteins (Steinmetz *et al.,* 2001), (c) building and testing of metabolic pathways (Ideker *et al.,* 2001) and (d) identification of uncharacterized genes and drug targets (Hughes *et al.,* 2000). These studies have shown that specific changes in the transcriptome or metabolome may occur due to gene deletion. The changes are expected to be more significant when the gene of a regulator protein is deleted.

2.10. Systematic Gene Deletions

Systematic gene deletion is a fundamental tool for functional genomics, since comparison of the metabolic state of cells in the presence and absence of a gene gives important clues about the function of that gene. Deletion mutant libraries including more than 6000 knock-outs were constructed after genome sequencing of *S. cerevisiae* was completed at 1996, and several methods are available to construct deletion mutants of *S.*

cerevisiae that employ deletion casettes (Güldener *et al.*, 1996, Giaever *et al.*, 2002). Mutant cells are grown in batch or chemostat culture under various environmental conditions and samples are analysed to determine the metabolic state of the cells.

2.10.1. Construction of Deletion Mutants

In the past, yeast researchers generated mutant collections through chemical or UV mutagenesis genetic screens. For instance, mutants of bc1 complex had to be identified among the broader class of respiratory-deficient nuclear petite (*pet*) mutants which were generated by chemical treatment of respiratory-competent strains, since no convenient procedures existed for selecting mutants of the bc1 complex (Tzagoloff, 1995). These procedures were often tedious and unsatisfactory; cloning and identification of all the affected genes could take years and it was impossible to determine if the screen was saturated. Recently, many novel approaches for the large-scale analysis of the yeast genome have been developed. These have led to the generation of valuable mutant collections that are commercially available and that can be screened by many researchers (Vidan and Snyder**,** 2001).

The precise deletion of yeast strains can be efficiently accomplished using a polymerase chain reaction (PCR)-mediated gene replacement strategy that exploits the high frequency and specificity of homologous recombination in yeast.

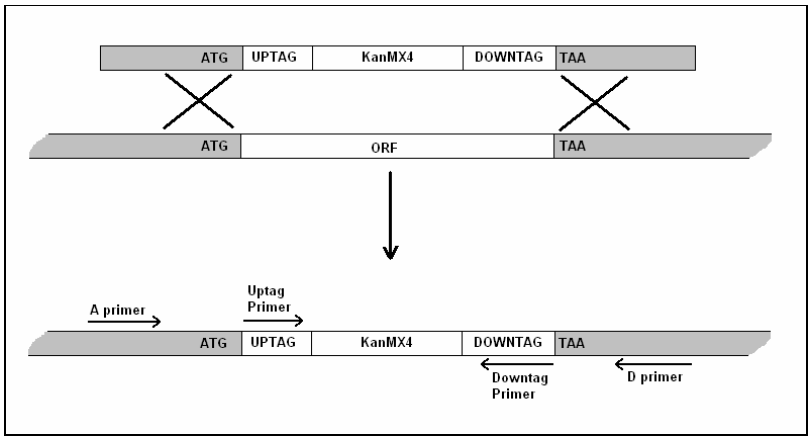

Figure 2.3. Deletion of ORFs by homozygous recombination (adapted from Saccharomyces Gene Deletion Project, 2002)

Güldener *et al.* (1996), have constructed the gene disruption cassette *loxP–kanMX–loxP*, which combines the advantages of the heterologous *kan*^{r} marker with those of the Cre–*lox*P system (Figure 2.3). The PCR-generated gene disruption cassette integrates with high efficiency via homologous recombination at the correct genomic locus (routinely 70%). In addition, the fast marker rescue procedure utilizing Cre-*lox*P system allows repeated use of the *kan*^{r} marker. Winzeler *et al.* (1999) reported that this method can be modified so as to introduce two molecular barcodes (UPTAG and DOWNTAG) into the deletion strain. The barcodes or "tags" are unique 20-base oligomer (20-mer) sequences that serve as strain identifiers.

The unique sequence tags linked to each gene deletion allow the strains to be analysed in parallel. In a competition experiment, a mixed culture containing every deletion mutant is grown, samples are collected at several times during growth, and the molecular bar code tags are amplified from genomic DNA. The abundance of each deletion strain is then determined by quantifying the associated molecular barcodes by hybridization to an oligonucleotide array of complementary barcode sequences. The more important a gene is for growth, the more rapidly the sequence tags of the corresponding deletion strain diminish in the culture. Thus, all genes required for growth can be identified and ranked in order of their relative contribution to fitness in a single experiment (*Giaever et al.*, 2002, Steinmetz *et al.*, 2002).

Winzeler *et al.* (1999) reported the construction of precise start-stop codon deletion mutants for 2026 *S. cerevisiae* ORFs. They used these strains in fitness screening (parallel competitive growth) experiments to analyse the genes essential for viability in yeast. They found that 1620 of the 2026 ORFs are not essential for viability, the essential ones are distributed evenly across the chromosomes and were more heavily transcribed. The protein products of essential genes which lack human homologs are thought to be the best targets for antifungal drugs.

A collection of over 11000 strains, each carrying a transposon inserted within a region of the genome expressed during vegetative growth and/or sporulation was generated by Ross-Macdonald *et al.* (1999). A multipurpose minitransposon (mTn), derived from bacterial transposible element Tn3, contains a *lacZ* reporter gene lacking an initiator

methionine and upstream promoter sequence. Introduction of this transposon into yeast results in production of β-galactosidase if the mTn is present within a transcribed and translated region of the genome, typicaly corresponding to an in-frame fusion of *lacZ* to yeast protein-coding sequence. Ross-Macdonald *et al.* (1999) have used the mutant collection, which affects nearly 2000 annotated genes, to determine disruption phenotypes for nearly 8000 strains using 20 different growth conditions; the resulting macroarray data sets were clustered to identify groups of functionally related genes.

Yan Tong *et al.* (2001) introduced the synthetic genetic array analysis (SGA) for construction of double deletion mutants claiming that inviable double-mutant meiotic progeny identify functional relationships between genes. Two genes show a "synthetic lethal" interaction if the combination of two mutations, neither lethal by itself, causes cell death. Synthetic lethal relationships may occur for genes acting in a single biochemical pathway or for genes within two distinct pathways if one of the genes compensates for or buffers the defects in the other. Pairs of such genes can be detected by SGA and mutants with multiple deletions can also be contructed to identify groups of interrelated genes.

The EUROSCARF (European *Saccharomyces Cerevisiae* Archive for Functional Analysis) collects yeast strains and plasmids that were generated during various yeast functional analysis projects. These include the following projects: BMBF (325 ORFs), EUROFAN I (800 ORFs) and EUROFAN II as part of the worldwide yeast gene deletion project (6000 ORFs). In a worldwide effort European and N. America laboratories are combined in the consortium of the Saccharomyces Genome Deletion Project (Winzeler *et al.*, 1999). The European task was part of EUROFAN II. All constructed strains were collected by EUROSCARF. The deletions were made in a genetic background isogenic to S288c, the *Saccharomyces cerevisiae* strain whose DNA sequence had been determined. Deletion strains constructed during phase II were all in the genetic background of the BY strain series (EUROSCARF, 2003).

2.10.2. Cultivation of Deletion Mutants

In genetic research, batch and continuous bioreactors are used unless inclusion of a new nutrient is needed for purposes like induction or labeling (e.g. for NMR spectrometric

analysis). In batch cultivation, the nutrients and cells to be grown are fed to the bioreactor at once. Samples are almost always collected at so called “mid-exponential” phase, when the cell growth rate is at its maximum. Samples may also be taken from the reactor at certain time intervals, if time dependent growth characteristics of the culture is of particular interest (e.g. DeRisi *et al.*, 1997). Continuous bioreactors are generally called ‘chemostats’ and enable balanced growth of the cells. While feed is supplied continuously to the reactor and homogenous mixing is done by an impeller or gas bubbles; exhausted growth medium including the cells, products and wastes is removed from the reactor at constant flow rate, so that the liquid amount in the reactor remains constant. After 5-10 recycles of growth medium, the cultivation reaches the steady-state or balanced growth, the culture removed from the reactor can be analysed as the steady-state product. For the cases that induction, labeling or other interference to growth is needed, fed-batch cultivation is employed. For any cultivation; temperature, pH and aeration should be under control and all the conditions should be well defined to ensure reproducibility of the results (Bailey and Ollis, 1994).

Baganz *et al.* (1997), constructed two different standard strains in which the *HO* gene is replaced by either *HIS3* or *kan*MX. They reported that there is a significant marker effect associated with *HIS3* which, moreover, is dependent on the physiological conditions used for the competition experiments. In contrast, the *kan*MX marker exhibited only a small effect on specific growth rate (<4 per cent). They suggested that nutritional markers should not be used to generate deletion mutants for the quantitative analysis of gene function in yeast but that *kan*MX replacements may be used, with confidence, for such studies. *HO* was chosen because it was found to be a selectively neutral site for gene replacement, it has no known role apart from mating-type swiching (inactive in diploids) and it has been used as the site of insertion of heterologous genes in brewing yeasts without any perceptible effect on the fermentation characteristics of the organism or quality of the product (Oliver *et al.*, 1998).

2.11. Transcriptome

DNA microarrays provide a format for the simultaneous measurement of the expression level of thousands of genes in a single hybridization assay. Each array consists

of a reproducible pattern of thousands of different DNAs (primarily PCR products or oligonucleotides) attached to a solid support, usually glass. Fluorescently labeled RNA or DNA prepared from messenger RNA is hybridized to complementary DNA on the array and then detected by laser scanning. Hybridization intensities for each DNA sequence on the array are determined using an automated process and converted to a quantitative read-out of relative gene expression levels. The transcriptome data can then be further analyzed to identify expression patterns and variations that correlate with cellular development, physiology and function (Harrington et al., 2000).

Several methods have been described for producing microarrays and two basic types of array technology are: spotted microarrays in which pre-synthesized single-strand DNAs are bound onto glass slides and high-density oligonucleotide arrays in which sets of oligomers are synthesized *in situ* on glass or silicon wafers using a photolithographic manufacturing process. High-density oligonucleotide arrays are available commercially. Although comprehensive expression profiles can be produced by both types of array, there are some fundamental differences between the two approaches (Figure 2.4).

On spotted arrays, genes are generally represented by single DNA fragments, greater than several hundred base pairs in length. The DNA samples hybridized to the array are, in most cases, labeled by incorporating fluorescently tagged nucleotides during oligo-primed reverse transcription of messenger RNA. Different fluorophores (generally Cy3- and Cy5-dUTP) are used to label cDNAs from control (reference) and experimental (test) RNAs. The labeled cDNAs are then mixed together prior to hybridization to the array. Relative amounts of a particular gene transcript in the two samples are determined by measuring the signal intensities detected for both fluorophores and calculating signal ratios.

On GeneChip (Affymetrix) oligonucleotide arrays, a given gene is currently represented by 15–20 different 25-mer oligonucleotides that serve as unique, sequence-specific detectors. In the GeneChip expression assay, eukaryotic mRNA is converted to biotinylated cRNA from oligo-dT-primed cDNA. Each sample is hybridized to a separate array. Transcript levels are calculated by reference to cRNA spikes of known concentration added to the hybridization mixture. Differences in mRNA levels between samples are determined by comparison of any two hybridization patterns produced on separate arrays of the same array type.

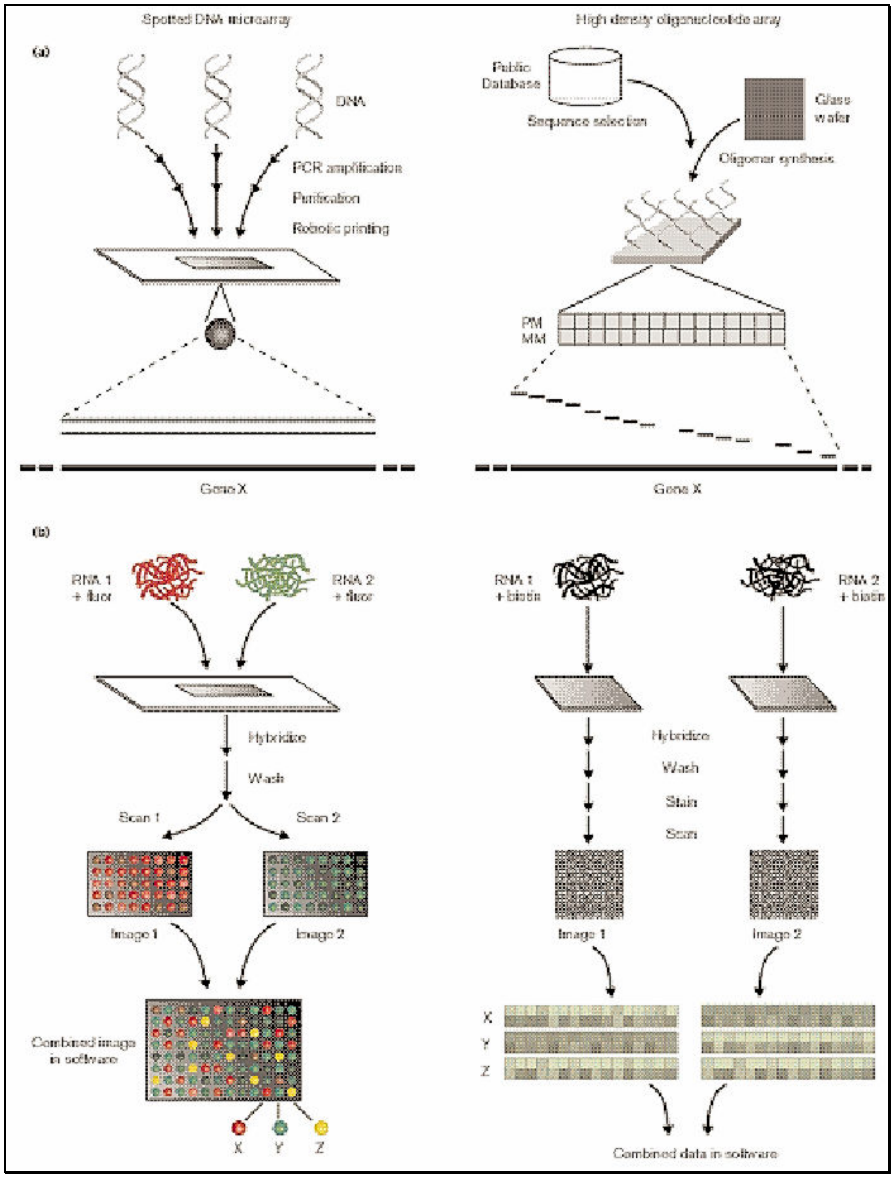

Figure 2.4. Spotted and oligonucleotide microarrays (Harrington *et al.*, 2000)

There are important differences in performance and information generated by the two types of arrays due to these differences in assay and gene representation. First, spotted arrays hybridized simultaneously with two distinctly labeled samples intrinsically normalize for noise and background in a pairwise comparison. The transcriptional read-out for these paired samples is provided as expression ratios and requires that different samples of an experimental set be hybridized with the same control or reference sample. The high-density oligonucleotide array assay, on the other hand, allows flexibility in sample comparisons and provides an estimate of the levels of gene transcripts in individual

samples. Second, on oligonucleotide arrays, the oligomer probes are designed to uniquely represent the cognate gene thus minimizing cross-hybridization between similar sequences. The potential for cross-hybridization between genes with significant levels of sequence similarity is high using spotted microarrays. Finally oligonucleotide arrays require gene sequence information for specifying the *de novo* synthesis of the oligomers on the array, whereas spotted arrays can be produced from both known and unknown cDNA and PCR fragments.

2.11.1. Transcriptome Analysis Studies on *S. cerevisiae*

The relatively small genome of yeast is well suited to global expression profiling with microarray technologies currently available. The availability of the complete genome sequence and well-characterized genetics has made *Saccharomyces cerevisiae* a popular focus of study with both spotted and oligonucleotide arrays. All ORFs and even intergenic regions can be represented, allowing unbiased assessment of an organism's expression profile.

First genome-wide mRNA expression measurements on spotted arrays were done by DeRisi *et al.* (1997) to investigate the temporal program of gene expression accompanying the metabolic shift from fermentation to respiration in *S. cerevisiae*. The genes with similar profiles during the progression of diauxic shift were known to share same binding motifs. Several classes of genes such as cytochrome c-related genes and those involved in TCA cycle and carbohydrate storage were found to be coordinately induced by glucose exhaustion. In another study (DeRisi *et al.*, 2000), identification of new targets of PDR transcription factors in various yeast deletion mutants was done by carrying out hybridization analysis on DNA microarrays.

Hughes *et al.* (2000) conducted mRNA level measurement experiments on 300 diverse mutations or chemical treatments in *S. cerevisiae* under a single growth condition and constructed a database or "compendium" of expression profiles using their data. They showed that cellular pathways affected from the perturbations can be determined by pattern matching, even among very subtle perturbations. They identified and experimentally

confirmed that eight uncharacterized ORFs encode proteins required for sterol metabolism, cell wall function, mitochondrial respiration or protein synthesis.

Ideker *et al.* (2001) analysed the GAL pathway in *S. cerevisiae* by using DNA microarrays, quantitative proteomics and databases of known DNA-protein and protein-protein interactions. Deletion strains of genes that are involved in yeast galactose-utilization pathway were grown in the presence and absence of galactose, 997 mRNAs were found to respond to these perturbations and 15 of 289 detected proteins are found to be regulated post-transcriptionally. This approach proved to be useful in testing and refining the models for metabolic pathways. In a separate study, effects of fermentable (galactose) and non-fermentable (ethanol) nutrients on mRNA and protein profiles of *S. cerevisiae* were investigated (Griffin *et al.*, 2002).

Transcriptional responses to different mitochondrial perturbations in *S. cerevisiae* were characterized by Epstein *et al.* (2001). Respiratory deficient petite cells and respiratory competent wild-type cells treated with inhibitors of oxidative phosphorylation were examined and the results showed that complete loss of the tricarboxylic acid cycle due to respiratory deficiency results in induction of a suite of genes associated with both peroxisomal activities and metabolite-restoration (anaplerotic) pathways.

Giaever *et al.* (2002), used deletion strains, each carrying a unique barcode that enable their abundance to be detected on DNA microarray assays. The whole collection of deletion strains of *S. cerevisiae* were grown under six conditions: high salt, sorbitol, galactose, pH 8, minimal medium and nystatin treatment to test the hypothesis that if the expression of a gene exhibits a significant increase at a given condition, then it should also be required for optimal growth in that condition. It was observed that some genes which are required for growth under a particular condition do not exhibit a change in expression at that condition, and it was concluded that the response to the change in conditions may operate post-transcriptionally.

The studies summarized above prove that DNA microarrays are extremely useful in determining the mRNA expression levels. When array experiments are coupled with well-controlled experimental systems and well-characterized genotypes, a striking amount of

highly interpretable information can be generated. The information can be used to assign function to unknown genes, enlarge understanding of cellular processes, identify potential drug targets and generate genome-wide snapshots of transcriptional activity in response to any stimulus or developmental trigger.

2.12. Metabolome Analysis

Metabolites are produced at the last step of the cellular machinery and define the biochemical phenotype of a cell or tissue. Quantitative and qualitative measurements of large numbers of metabolites thus provide a broad view of the biochemical status of an organism that can be used to monitor and assess gene function. Profiling of the transcriptome and proteome has received some criticism due to its inability to predict gene function in some cases. Although the transcriptome represents the delivery mechanism of a translational code to the cellular machinery for protein synthesis, increases in mRNA levels do not always correlate with increases in protein levels (Gygi *et al.*, 1999). Furthermore, once translated, a protein may or may not be enzymatically active. Due to these factors, changes in the transcriptome or in the proteome do not always correspond to alterations in biochemical (i.e. metabolic) phenotypes. Another consideration when profiling the transcriptome and proteome is that most modern techniques identify mRNA and proteins through sequence similarity or database matching; thus, identification is based primarily on the quality of the match and is therefore indirect. In the absence of existing database information, transcript or protein profiling often yields only limited information. Based on the above limitations, profiling the metabolome may actually provide the most ‘‘functional’’ information of the ‘‘omics’’ technologies. There are, nevertheless, many instances in which transcriptome and proteome profiling have successfully pointed the observer to functional information, and therefore an integrated approach is preferred when resources permit (Sumner *et al.*, 2003).

For a number of metabolites, kits for concentration measurement using enzymatic methods are commercially available. Though the results obtained are reliable and concentrations are obtained readily without utilizing any software, these methods are time-consuming when a large number of samples is to be handled, moreover kits are available for a limited number of metabolites like glucose, ethanol, and a few acids like succinate

and pyruvate. The kits are suitable and satisfactory if a low number of samples is to be analysed with the aim of determining basic metabolome data like amount of glucose utilized or ethanol produced.

Since the techniques for obtaining mRNA and protein profiles allow thousands of species to be detected in one reaction, it is desired to have the corresponding metabolome data to follow the cell machinery until the last step. It is possible to screen for a high number of metabolites simultaneously using analytical tools such as chromatography and spectroscopy. Methods based on infrared spectroscopy (IR), nuclear magnetic resonance (NMR), thin-layer chromatography (TLC), HPLC with ultraviolet and photodiode array detection (LC/UV/PDA), capillary electrophoresis coupled to ultraviolet absorbance detection (CE/UV), capillary electrophoresis coupled to laser-induced fluorescence detection (CE/LIF), capillary electrophoresis coupled to mass spectrometry (CE/MS), gas chromatography-mass spectrometry (GC/MS), liquid chromatography-mass spectrometry, (LC/MS), liquid chromatography tandem mass spectrometry (LC/MS/ MS), HPLC coupled with both mass spectrometry and nuclear magnetic resonance detection (LC/NMR/ MS) and LC/NMR/MS/MS have all been used to measure metabolites (Sumner *et al.* 2003).

The selection of the most suitable technology is generally a compromise between speed, selectivity and sensitivity. Tools such as NMR are rapid and selective, but have relatively low sensitivity. Other methods such as capillary electrophoresis coupled to laser induced fluorescence (CE/LIF) detection are highly sensitive, but lack selectivity. Hyphenated mass spectrometry methods such as GC/MS and LC/MS offer good sensitivity and selectivity, but relatively longer analysis times (Sumner *et al.* 2003).

2.12.1. Metabolome Analysis Studies on *S. cerevisiae*

A small number of extracellular and intracellular metabolites of *S. cerevisiae* have been quantified by many research groups using enzymatic methods with aim of defining the metabolic state of yeast under various conditions. These limited data were obtained by enzymatic methods, HPLC or NMR and frequently used for the kinetic modelling of central metabolism. However, in the post-genomic era, metabolome data are expected to be

a powerful tool for functional analysis. Only two recent studies will be reviewed below, with the aim of illustrating the importance of metabolome data.

A functional analysis study was done by Raamsdonk *et al.* (2001) that used metabolic fingerprinting to reveal the phenotype of silent mutations of *S. cerevisiae*. The term "metabolic fingerprinting" corresponds to determination of the intracellular metabolome. Comprehensive analysis of metabolites within yeast cell extracts was done using FTIR, ES-MS and NMR spectroscopy. Unfortunately, measuring intracellular metabolites is time-consuming and subject to technical difficulties caused by the rapid turnover of intracellular metabolites and need to quench metabolism and separate metabolites from the extracellular space. Thus, this method is difficult to scale-up for high-throughput screening.

Metabolic footprinting focuses on direct mass spectrometric monitoring of extracellular metabolites in spent culture medium. Allen *et al.* (2003) used metabolic footprinting to distinguish between different physiological states of wild-type yeast and between yeast single-deletion mutants even from related areas of metabolism (Figure 2.5). Using principle componenets analysis (PCA) and genetic programming, Allen *et al.* (2003) showed that metabolic footprinting is an effective method to classify unknown gene deletions. Footprinting was reviewed recently with all aspects (Kell *et al.,* 2005).

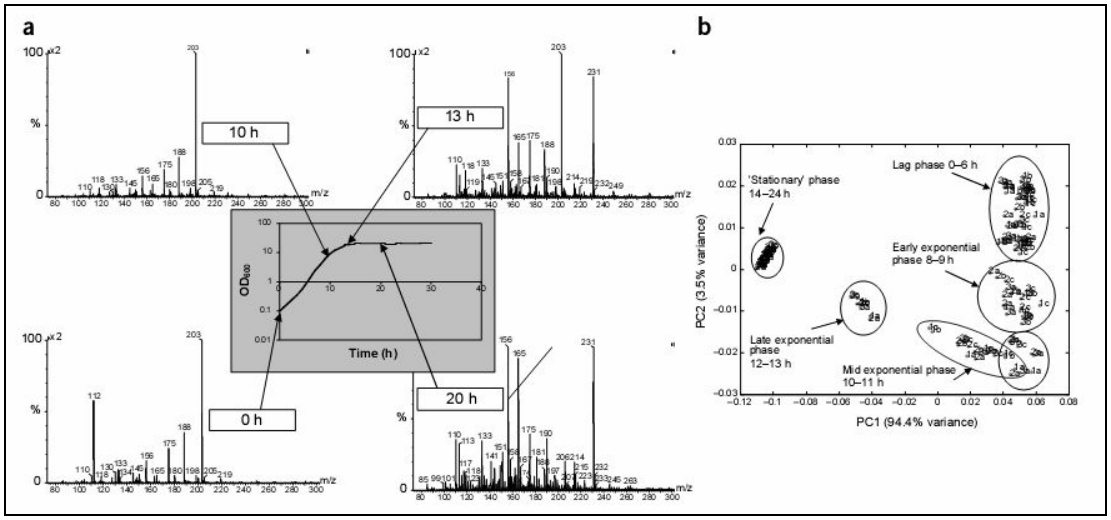

Figure 2.5. Metabolic footprinting of *S. cerevisiae*, a) Representative spectra, b) PCA of data (*Allen et al*., 2003)

Most of the genes involved in fermentation and respiration by *S. cerevisiae* are known. However, there are unclear points in the regulation of the pathways that are part of fermentation and respiration. Metabolic footprinting and fingerprinting of fermenting or respiring yeast might be used to identify new genes involved in these pathways. Moreover, the extent of fermentation and respiration in an uncharacterized culture might be quantified by comparing its profile with profiles of previous cultivations make conclusions about the physiology of the culture in question.

2.13. Functional Annotation of Genes

Following the complete sequencing of genomes, attempts were made (for many organisms) to annotate functions to the unknown genes depending on the similarity of their sequences with the sequences of known genes (Cho *et al*., 2001; Curwen *et al*., 2004; Mills *et al*., 2003; Zhou *et al*., 2002; Khan *et al.*, 2003; Koski *et al*. 2005; Martin *et al*. 2004; Vinayagam *et al.* 2004). Collection of high-througput experimental data on protein-protein or protein-DNA interactions enabled new annotations depending on interaction networks (Deng et al. 2004; Brun 2004, Letowski and Kasif 2003; Zhou et al. 2002a). Text mining in web resources and available publications were also used for annotation of new functions (Tamames *et al.*, 2005).

As the experimental data in databases accumulated, integration of different data sets became possible for annotations depending on many biological evidences (Schlitt et al 2003; Chen 2004; Hazbun 2003; Troyanskaya et al. 2004). The database STRING gathering all these information together (neighborhood on the genome, fusion information from other species, phylogenetic profile, coexpression, protein-protein interaction (PPI) and protein complexes, information from databases, textmining in abstracts of publications) is used to investigate all functional linkages at a glance (von Mering *et al.,* 2005) and also bioinformatics tools developed for integrating all biological data is released for public use (*e.g.* AVID by Jiang and Keatling (2005); GeneFAS by Joshi *et al.,* (2004)).

In Table 2.4, a list of publications reporting methods for *in silico* functional genomics of *S. cerevisiae* and Gene Ontology (GO) is given. This list is far from being

complete, however indicates the trends of computational functional genomics research. This section concludes the survey of historical and recent research on *S. cerevisiae*.

Table 2.4. Recent Publications on *In Silico* Functional Genomics

Publication	Title	Classification	Data Type
Cho and Walbot, 2001	Computational methods for gene annotation: the *Arabidopsis* genome	GenBank, Arabidopsis Genome Initiative	Sequence Homology
Zhou *et al*., 2002	UniBLAST: A system to filter, cluster, and display BLAST results and assign unique gene annotation	-	Sequence Homology
Zhou *et al*., 2002b	Transitive functional annotation by shortest-path analysis of gene expression data	GO (SGD)	Expression correlation, Co-localization
Cornell *et al*., 2003	GIMS: an integrated data storage and analysis environment for genomic and functional data	GO	Transcriptome, PPI, genomic data, metabolic pathways
Hazbun *et al*., 2003	Assigning Function to Yeast Proteins By Integration of Technologies	GO	PPI, Co-localization, co-purification, homology and structural similarity
Khan *et al*., 2003	GoFigure: Automated Gene Ontology annotation	GO (SGD)	Sequence Homology
King et al., 2003	Predicting phenotype from patterns of annotation	GO (SGD)	Yeast Phenotype Catalog
Letowsky and Kasif 2003	Predicting protein function from protein/protein interaction data: a probabilistic approach	GO (SGD)	PPI

Table 2.4. – continued

Rougemont Hingamp, 2003	DNA microarray data and contextual analysis of correlation graphs	GO (SGD)	Gene Expression
Schlitt et al., 2003	From Gene Networks to Gene Function	GO (SGD)	PPI, Protein complexes, text search, gene expression, TF binding site
Troyanskaya *et al*., 2003	ABayesian framework for combining heterogeneous data sources for gene function prediction	GO (SGD)	PPI, Gene Expression, Co-localization, TF binding
Alloco et al., 2004	Quantifying the relationship between co-expression, co-regulation and gene function	GO (SGD)	Gene Expression, TF binding
Brun et al., 2004	Clustering proteins from interaction networks for the prediction of cellular functions	YPD classificatio n and GO (SGD)	PPI
Chen and Xu, 2004	Global protein function annotation through mining genome-scale data in yeast Saccharomyces cerevisiae	GO (SGD)	PPI, protein complexes, gene expression, sequence homology, co-localization
Curwen et al., 2004	The Ensembl Automatic Gene Annotation System	SWISS-PROT and RefSeq	Sequence Homology
Deng et al., 2004	Mapping gene ontology to proteins based on protein–protein interaction data	GO (SGD)	PPI

Table 2.4. – continued

Joshi et al., 2004	Genome-Scale Gene Function Prediction Using Multiple Sources of High-Throughput Data in Yeast Saccharomyces cerevisiae	GO (SGD)	PPI, Protein complexes, Phenotypes, Protein Classes and Motifs, Gene Expression
Martin et al., 2004	GOtcha: a new method for prediction of protein function assessed by the annotation of seven genomes	GO	Sequence Homology
Vinaygam et al., 2004	Applying Support Vector Machines for Gene ontology based gene function prediction	GO	Sequence Homology
Jiang and Keatling, 2005	AVID: An integrative framework for discovering functional relationships among proteins	GO (SGD)	PPI, Sequence Homology, Protein Complexes, Localization, Gene Expression
Koski et al., 2005	AutoFACT: An Automatic Functional Annotation and Classification Tool	Metabolic pathways, Functional classes, Enzyme classes, GO	Sequence Homology
Nabieva et al., 2005	Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps	MIPS functional hierarcy and GO (SGD)	PPI, co-expression and shared evaluationary information
Rubinstein and Simon, 2005	MILANO – custom annotation of microarray results using automatic literature searches	User defined terms	Text search in web

Table 2.4. – continued

Tamames, 2005	Text Detective: a rule-based system for gene annotation in biomedical texts	-	Text mining
Valencia, 2005	Automatic annotation of protein function	-	Discussion on function annotation methods

3. MATERIALS AND METHODS

Materials and methods will be given in two subsections, the experimental part lists the equipment and materials, and summarizes the protocols used in the experiments. The computational part gives a list of the software tools and summarizes the computational methods.

3.1. Experimental Materials and Methods

This section gives the details of the strains, chemicals, equipments, analysis kits used in this study followed by the experimental procedures applied.

3.1.1 Experimental Materials

3.1.1.1. *S. cerevisiae* strains. Deletion strains of *S. cerevisiae* with genomic background *BY4743* (*MAT**a**/α his3*Δ1/*his3*Δ1 *leu2*Δ0/*leu2*Δ0 *lys2*Δ0/*LYS2 MET15*/*met15*Δ0 *ura3*Δ0/*ura3*Δ0) from the Yeast Genome Deletion Project library were used. The *ho* deletant is commonly used as a standard strain in control experiments since the deletion has no measurable impact on either flux (growth rate; Baganz *et al*.,1997) or metabolome (Oliver *et al*., 1998). The absence of the genes in a strain's genome was verified using PCR-based methods. A list of deletion strains used in this study is given in Table 3.1.

3.1.2.2. Primers. Primers were designed to amplify the deletion casettes for PCR-based verification of gene deletions. The list of the primers are given in Table 3.2. The sequences provided by Yeast Genome Deletion Project are used. Locations of the primers on the genome are given in Figure 3.1. KanMX forward and reverse primers were used to verify the existence of KanMX cassette on the genome. Uptag-Downtag primers are used to verify the existence of correct deletion cassette and A-D primers are used to verify that deletion cassette is in correct location on the genome.

Table 3.1. Deletion mutants

hoΔ/hoΔ	Homozygous deletion mutant of BY4743
hap4Δ/hap4Δ	Homozygous deletion mutant of BY4743
oxa1Δ/oxa1Δ	Homozygous deletion mutant of BY4743
bcs1Δ/bcs1Δ	Homozygous deletion mutant of BY4743
rip1Δ/rip1Δ	Homozygous deletion mutant of BY4743
mig1Δ/mig1Δ	Homozygous deletion mutant of BY4743
mba1Δ/mba1Δ	Homozygous deletion mutant of BY4743
HAP4/hap4Δ	Heterozygous deletion mutant of BY4743
RIP1/rip1Δ	Heterozygous deletion mutant of BY4743
MIG1/mig1Δ	Heterozygous deletion mutant of BY4743

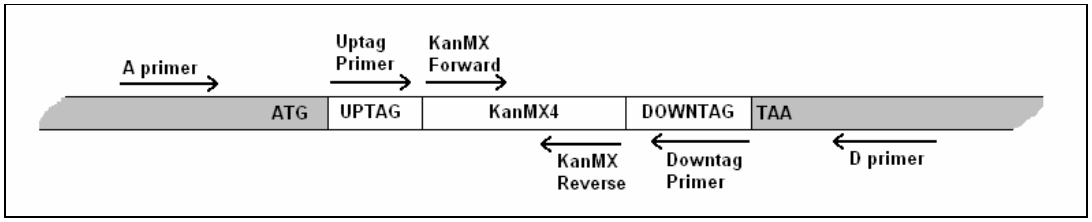

Figure 3.1. Locations of the primers on deletion locus (adapted from Yeast Genome Deletion Project, 2002)

3.1.2.3. Isolation and Assay Kits. Isolation of genomics DNA was carried out using TAKARA DNA isolation kit (TAKARA Bio Inc., Catalog number: 9084). Ethanol and glucose contents of the samples were measured using Boehringer Mannheim/R-Biopharm analysis kits (Catalog numbers: 10 176 290 035 and 10 716 251 035).

Table 3.2. Primers and Sequences

Gene	Primer	Sequence
-	KanMX.Forward	CTGCAGCGAGGAGCCGTAAT
-	KanMX.Reverse	TGATTTTGATGACGAGCGTAAT
BCS1	YDR375C.DOWNTAG	AATTAGATGGGACGTTGCCT
BCS1	YDR375C.UPTAG	AAGTCCGCTTAACCTATGCG
RIP1	YEL024W.DOWNTAG	CAGCGCATCCAATTACATGA

Table 3.2. continued

RIP1	YEL024W.UPTAG	CTATTCTGGAGCACCGATGG
MIG1	YGL035C.DOWNTAG	CTTAGTGCTAGAGTTACGGT
MIG1	YGL035C.UPTAG	CCACGAGACGTTTCTAAATC
HAP4	YKL109W.DOWNTAG	CCATAAGACGTTGTGAAACC
HAP4	YKL109W.UPTAG	TGGAATACCCACGAACGGAC
HO	YDL227C.DOWNTAG	GCGCGTCAAGTACAAAGTTA
HO	YDL227C.UPTAG	GGGCACACCCAATTTAGACA
OXA1	YER154W.DOWNTAG	TAAACATAGCGCCAGGACGG
OXA1	YER154W.UPTAG	ATCTGTGCGCCAGTATCGCT
BCS1	YDR375C.A	GTTTATTTGTGAAGAAATGT
BCS1	YDR375C.D	TTAGTTTAGATGCGGAGGAG
RIP1	YEL024W.A	TATTTCATCCTTTCAACTTC
RIP1	YEL024W.D	GAAAAAGAAGATGGTGAGAC
MIG1	YGL035C.A	GAAGCAACAACAAATTTTTA
MIG1	YGL035C.D	GAACAATTAATTATCTCTGC
MBA1	YBR185C.A	GGAAGAGTCTCAGATATTGG
MBA1	YBR185C.D	TCTAAAAGATGAAGCAAAGC
OXA1	YER154W.A	AGTCTTTTCGTGATAGTCAA
OXA1	YERI54W.D	CATGCAATGGTAAATGAAGA
OXA1	YER154W.UPTAG	ATCTGTGCGCCAGTATCGCT
OXA1	YER154W.DOWNTAG	TAAACATAGCGCCAGGACGG

3.1.2.4. Microarrays. The GeneChip Yeast Genome S98 Array (YG-S98) used in this study is a single array that permits the monitoring of the mRNA abundance from approximately 6,400 open reading frames (ORFs) and additional sequences from the yeast genome. Oligonucleotide probes were synthesized on the array complementary to a portion of each gene represented. The genes represented on the array were selected predominantly from the Saccharomyces Genome Database (SGD, Dolinski *et al*., 2005) with a download date of 12/98. Additional sequences were obtained from the Munich Information Center for Protein Sequence (MIPS, Mewes *et al*., 2004).

Also probes for potentially-expressed sequences identified by Serial Analysis of Gene Expression were included. On YG-S98, there are approximately 16 pairs of probes for each gene ("*Affymetrix GeneChip Expression Analysis Technical Manual*", 2000-2003, Affymetrix Inc).

The GeneChip probe array consists of a square glass substrate mounted in a plastic cartridge (Figure 3.2). The glass contains an array of oligonucleotides that, when mounted, is on the inner glass surface. A chamber in the plastic housing directly under the glass acts as a reservoir where hybridization and washing occur ("*Affymetrix GeneChip Expression Analysis Technical Manual*", 2000-2003, Affymetrix Inc).

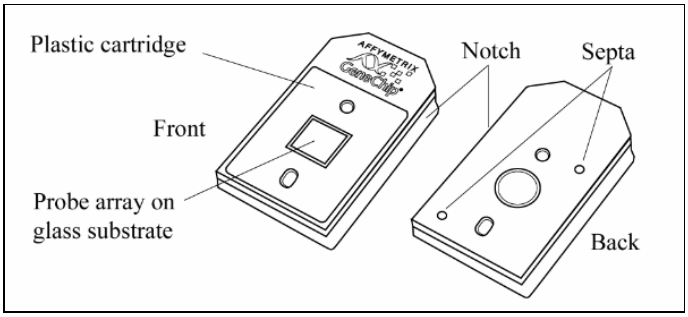

Figure 3.2. GeneChip Probe Array (Affymetrix, 2000-2003)

3.1.2.5.Chemicals. Chemicals used in this study are listed in Table 3.3.

Table 3.3. Chemicals and suppliers

Chemical	Supplier
Acotonitrile	Merck
Agar-agar	Merck
Ammonium sulfate - $(NH_4)_2SO_4$	Merck
Biotin	Merck
Boric Acid	Merck
Boric Acid - H_3BO_3	Merck
Bromophenol blue	Merck
Calcium chloride - $CaCl_2$**.** $2H_2O$	Merck
Ca-pantothenate	Merck

Table 3.3. continued

Chloroform	Merck
Cupper (II) sulfate pentahydrate - $CuSO_4 . 5H_2O$	Merck
DEPC (diethylpyrocarbonate)	Sigma
EDTA (ethylenediaminetetraacetic acid)	Merck
Ethanol	Merck
Ethidium bromide	Sigma
Ferric chloride hexahydrate - $FeCl_3 . 6H_2O$	BDH
Formic acid	Merck
G-418 (Geneticin)	Sigma
Glucose	Merck
HEPES	Sigma
Inositol	Merck
IPA (isopropyl alcohol)	Merck
L-histidine	Merck
Lithium chloride - LiCl	Merck
L-leucine	Sigma
Magnesium sulfate heptahydrate - $MgSO_4 . 7H_2O$	Merck
Methanol	Merck
NTPs	Promega
Peptone	Merck
Potassium dihydrogen phosphate - KH_2PO_4	Merck
Potassium hydroxide - KOH	Merck
Potassium iodide - KI	Merck
Pyridoxine	Fluka
Sodium Chloride - NaCl	Merck
Sodium chloride - NaCl	Merck
Taq DNA polymerase	Promega
Thiamine/HCl	Merck
TRIZOL	Invitrogen
Tris/HCl	Merck
TrisAcetate	Merck

Table 3.3. continued

TrisBase	Merck
Uracil	Fluka
Yeast extract	Merck
Zinc sulfate heptahydrate - $ZnSO_4 \cdot 7H_2O$	Merck

3.1.2.6. Laboratory Equipment. List of laboratory equipment and suppliers are given in Table 3.4.

Table 3.4. Laboratory equipment

Equipment	Supplier
Autoclave	Eyela Model MAC-601(Japan)
Balance	Precisa 80A-200M (Switzerland)
Centrifuges	SORVALL RC-5B Refrigerated Superspeed Centrifuge, DuPont (USA) Eppendorf 5415 C (Germany)
Deep freezers	-80^{o}C, Hetofrig CL410HETO (Denmark) -20^{o}C, BOSCH (Germany)
Dismembrator	Biolab Micro-Dismembrator S (New England, USA)
Horizontal gel-electrophoresis	Horizon 58, Model 200, BRL (USA)
Fermenters	Bioflo III Batch/Continuous Fermenter, New Brunswick (England)
Fludics Equipment	Affymetrix (USA)
Heating Magnetic Stirrers	MR 3001, Heidolph (Germany) Scientifca ARE, VELP (Italy)
Hybridization oven	Affymetrix (USA)
Incubator	NÜVE EN500 (Turkey)
Laminar Flow Cabinet	HBB 2460 LaminAir, Holten (Denmark)
Orbital Shakers	GFL 3032, GFL (Germany) INNOVA 4340 Illuminated Refrigerated Incubator Shaker, New Brunswick Scientific (USA)

Table 3.4. continued

Refrigerators	+4°C Ariston (Italy) +4°C Arçelik (Turkey)
Rotavapor	HETO VR1 (Denmark)
Scanner	Affymetrix (USA)
Spectrophotometer	DU 640 Beckman (USA)
Vortex	Elektromag (Turkey)
Water Baths	HETO, CB 8-30e AT_{110} (Denmark) HETO, CB 8-30e DT_1 (Denmark) HETO DT Hetotherm (Denmark)
Water Purification Systems	Millipore, Milli Ro Plus (USA) Millipore, Milli-Q UF Plus (USA)

3.1.3. Experimental Methods

The experimental methods used are described in detail in the following sections.

3.1.2.1. Plates. G418-containing (200 mg/l) YPDA medium (2 per cent yeast extract, 1 per cent peptone, 2 per cent glucose, 2 per cent agar-agar), was sterilized in an Erlenmeyer flask and transferred to sterile plates (30 ml portions) and allowed to cool and solidify. Cells from 10-20 microliters of frozen stock was distributed on the surface of the solidified medium, the plate was isolated using Parafilm and incubated at 30°C for 1 or 2 days until the colonies are visible. The plates with colonies were kept at 4°C. The plates for respiration tests were prepared similarly using G418-containing (200 mg/l) YPEA (2 per cent yeast extract, 1 per cent peptone, 2 per cent ethanol, 2 per cent agar-agar).

3.1.2.2. Preculture. 10 ml G418-containing (200 mg/l) YPD medium (2 per cent yeast extract, 1 per cent peptone, 2 per cent glucose, 2 per cent agar-agar) was sterilized in an Erlenmeyer flask, seeded with 1-2 colonies from the YPDA plates. The culture was incubated at 30°C overnight.

3.1.2.3. Media preparation. Mineral media, supplemented with trace elements and vitamins were used (Baganz *et al.*, 1997) in chemostat experiments. The compositions of the media

were as follows: KH_2PO_4 (2 g/l), $MgSO_4.7H_20$ (0.55 g/l), NaCl (0.1 g/l), $CaCl_2.2H_2O$ (0.09 g/l), Uracil (0.02 g/l), L-Histidine (0.02 g/l), L-Leucine (0.1 g/l), $ZnSO_4.7H_2O$ (0.7×10^{-4} g/l), $CuSO_4.5H_2O$ (0.1×10^{-4} g/l), H_3BO_3 (0.1×10^{-4} g/l), KI (0.1×10^{-4} g/l), $FeCl_3.6H_2O$ (0.5×10^{-4} g/l), inositol (0.12 g/l), thiamine/HCl (0.014 g/l), pyridoxine (0.004 g/l), Ca-pantothenate (0.004 g/l), biotin (0.0003 g/l). For glucose-limited medium, 3.13 g/l $(NH_4)_2SO_4$ and 2.5 g/l glucose were added to the medium described above. For nitrogen-limited medium 0.46 g/l $(NH_4)_2SO_4$ and 20 g/l glucose were added to medium described above.

3.1.2.4. Fermentation. The chemostat reactors used in this study were BioFlo 3000 fermentors with in/out flow streams, agitation, temperature and pH control. The feed stream was supplied from a 20L container which was sterilized prior to operation and filled with filter-sterilized medium under sterile conditions. The fermentors were aerated with filter-sterilized air supplied from a compressor. An external heating belt connected to a controller was used to keep the fermentors at constant temperature. 1M KOH and 1M HCl solutions were used as basic and acidic supplies for adjusting the pH, flowrates of these solutions were adjusted by the controller.

Prior to operation, the pH probe was calibrated and fitted to fermentors. DO probe was fitted with no calibration. The fermentors were autoclaved and filled with medium. Connections for temperature and pH control were made, calibration of DO probe was performed by saturating the medium with O_2 (saturation point calibration) and with N_2 (zero point calibration). The level control bar was adjusted to keep the working volume at 1L. PID control of agitation (750 rpm), pH (4.5) and temperature (30^oC) were activated. Pumps were calibrated for the flowrate to be used (100 ml or 200 ml per hour for D = $0.1hr^{-1}$ and D = 0.2 hr^{-2}, respectively). Precultures (10 ml) grown overnight in G418-containing YPD medium were used to inoculate the fermentors. Temperature and pH of chemostats were kept constant at 30^oC and 4.5, respectively, and the oxygen content was maintained at saturation. The total operation time varied from 72 hours to 120 hours.

3.1.2.5. Sample collection. The biomass content of the fermentors were measured as OD (optical density) at 600 nm using acrylate cuvettes and a spectrophotometer. Steady operation was assumed to be reached when OD changes were no more than 10 per cent.

After three retention times (30 and 15 hours for D = 0.1hr^{-1} and D = 0.2 hr^{-2}, respectively), the samples were collected into precooled tubes and dipped into liquid nitrogen to freeze the samples immediately. 1 ml samples were collected for extracellular metabolite analysis (4 x 1ml), 5 ml samples were collected for intracellular metabolite analysis (2 x 5ml), 20 ml samples were collected for mRNA analysis (2 x 20 ml). The samples were kept at -80^{o}C.

3.1.2.6. RNA isolation. Isolation of total RNA was carried out using Trizol reagent. The 20 ml samples from fermentors were defrosted on ice and following steps were carried out:

a. Cells were harvested by centrifugation at 4000 rpm for 3 minutes.
b. 5ml teflon vessel was precooled in liquid nitrogen together with a 7mm bead made of tungsten carbide. Vessel was taken out of liquid nitrogen with the bead and some liquid nitrogen remaining in it.
c. Cell pellet was resuspended in a very small volume of growth medium.
d. Suspension was sucked into a pipette and released as individual drops directly into liquid nitrogen in the pre-cooled teflon vessel.
e. The vessel was closed with a pre-cooled cap and placed into the holder of micro-dismembrator.
f. The vessel was shaken at 1200rpm for 5 minutes.
g. The frozen powder was dissolved in 2 ml of Trizol and split into two eppendorfs.
h. The suspension was homogenized by vortexing for 1 min.
i. Samples were kept at room temperature for 5 min to allow for dissociation of the nucleoprotein complexes.
j. Samples were centrifuged at 12000g for 5 min and clear supernant was transfered to a fresh tube (this was an optional step to precipitate polysaccharides, extracellular matrix and high molecular weight DNA).
k. Samples were shaken vigourously for 15 sec after addition of 0.4 volume (400 microliters) chloroform and left at room temperature for 5 minutes.
l. Samples were centrifuged at 12000g for 5 min. The mixture separates into the lower red phase (phenol-chloroform), the interphase and the colourless upper aqueous phase. RNA was forced exclusively into the aqueous phase whereas the DNA and the proteins partition into the interphase and the lower phenol phase).

m. Aqueous phase was transferred to a fresh tube (0.5 ml). Aqueous phase from two identical sample tubes were combined to yield 1 ml of sample.
n. RNA was precipitated with 0.5 ml of wasopropil alcohol (IPA) / ml of sample and kept at room temperature for 5-15 minutes. Samples can be left overnight at this stage.
o. Samples were centrifuged at 12000g for 10 min. RNA pellet should form a gel like precipitate on the bottom and side of the tube.
p. IPA was removed carefully.
q. The pellet was washed with 1ml 70 per cent ethanol by vortexing and centrifuged at 12000g for 10 min.
r. RNA pellet was briefly air dried.
s. RNA was dissolved in 0.5ml DEPC treated water.
t. 1 volume of LiCl buffer was added and precipitated at least for one hour at -20^{o}C. Samples were centrifuged at maximum speed for 30 min. Samples can be left overnight at this stage.
u. Pellet was washed with 1 ml 70 per cent ethanol to remove any salts as these will act as inhibitors of reverse transcriptase.
v. Pellet was dried and dissolved in 25 microliters of DEPC treated water.
w. RNA concentration of 2:500 diluted RNA sample in DEPC treated water was determined by measuring OD at 260 nm. 1 microliter of samples were diluted 1:10 and run on 1 per cent agarose gel (TBE) to visualize and verify the quality of RNA.

3.1.2.7. DNA isolation. DNA isolation was carried out using TAKARA Dr.GenTLE DNA isolation kit (TAKARA Bio Inc., Catalog number: 9084). The steps applied were as follows:

a. Cell culture was centrifuged at 12000rpm for 30 sec.
b. Supernatant was discarded.
c. 540 microliter of GenTLE Yeast Solution I was added to pellet and mixed by vortexing.
d. 60 microliter of GenTLE Yeast Solution II was added and mixed by vortexing.
e. Mixture was heated at 70^{o}C for 10 min.
f. 300 microliter of GenTLE Yeast Solution III was added and mixed by converting the tube (no vortex).

g. The tube was placed on ice for 2 min and mixed again by converting the tube (no vortex).
h. Mixture was centrifuged at 12000 rpm for 5 min at 4^{o}C.
i. Supernatant was transferred to a fresh tube. 600 microliters of isopropanol was added and mixed by vortexing.
j. Mixture was centrifuged at 12000 rpm for 5 min at 4^{o}C, supernatant was discarded. White DNA pellet should be visible.
k. DNA pellet was washed with cold 70 per cent ethanol solution and centrifuged again at 12000 rpm for 5 min at 4^{o}C.
l. Supernatant was discarded and DNA pellet was dried either by air or vacuum.
m. Pellet was dissolved in 50 microliters of TE buffer.
n. Resuspended DNA was stored at 4^{o}C or -20^{o}C.

3.1.2.8. PCR mixture. The amplification of DNA samples were carried out using PCR reactions. The PCR reaction mixtures were prepared as follows:

a. Template DNA : 1.0 µl
b. 10 x Buffer : 5.0 µl
c. 10 mM NTPs : 0.4 µl
d. Primer 1 (1:50) : 2.5 µl
e. Primer 2 (1:50) : 2.5 µl
f. H_20 : 37.6 µl
g. Taq DNA Polymerase: 1.0 µl

3.1.2.9. PCR program. The PCR program used to amplify the DNA from deletion mutants was as follows:

a. 94^{o}C - 3 min - 1 cycle
b. 94^{o}C - 15 sec
   55^{o}C - 15 sec
   72^{o}C - 1 min
c. The section (b) applies for 35 cycles
d. 72^{o}C - 3 min - 1 cycle
e. 4^{o}C - storage period

3.1.2.10. Gel electrophoresis. Gel electrophoresis was used to visualize DNA and RNA samples. 0.7 per cent agarose gel in TAE buffer was prepared and samples were visualized as follows:

a. 0.7 gr agarose was dissolved in 100 ml TAE buffer.
b. The solution was heated until the boiling point in microwave oven.
c. Solution was cooled at room temperature for 5 min.
d. 4 ml of ethydium bromide solution was added and mixed.
e. The gel was poured on the casting tray and allowed to cool.
f. The gel tank was filled with TAE buffer (gel should be covered with buffer)
g. 9 μl of DNA or RNA sample was mixed with 1 μl loading dye.
h. The mixture was loaded to well in the gel, they are run for 45 min under 120V.
i. The bands were visualized under UV light.
j. 1 per cent agarose gel in TBE buffer was prepared similarly using 1 gr of agarose and replacing TAE buffer with TBE buffer.

3.1.2.11. Intracellular Metabolite Extraction. The method applied for intracellular metabolite extraction was adapted from Gonzales *et al.* (1997) and Castrillo *et al.* (2003):

a. 5 ml cell cultures were quenched in 20 ml 60 per cent methanol, 70mM HEPES, 7.5 pH solution at -40^{o}C.
b. The mixture was stored at -80^{o}C.
c. The mixture was centrifuged at 12000 rpm for 3 min.
d. Metabolites from the pellet were extracted by adding 5 ml of boiling 75 per cent ethanol 0.25M HEPES, 7.5 pH solution.
e. The samples were analysed for glucose content using enzymatic method.
f. The samples were analysed for metabolome profile using MS.

3.1.2.12. Enzymatic Determination of Glucose Content. Boehringer Mannheim/R-Biopharm analysis kit (Catalog number: 10 716 251 035) was used for analysis of glucose content in samples. The procedure was as follows:

a. 1 ml solution 1 was pipetted into cuvettes.
b. 0.1ml of sample was pipetted into the cuvettes.
c. 1.9 ml of redistilled water was pipetted into the cuvettes.
d. The mixture was mixed by pipetting or converting the cuvette .

e. After approximately 3min, absorbance of the mixture was measured at 340 nm using spectrophotometer (A_1).
f. 0.02 ml of suspension 2 was added to the mixture, mixed by pipetting.
g. After completion of reaction (approx. 10-20 min), the absorbances were read again (A_2).
h. The same procedure was applied for a standard reaction, 0.1ml of water replaced the sample in the standard reaction.
i. Glucose contents of the samples were calculated as described in the manual of the kit. The samples were be diluted prior to analysis when the glucose concentration was higher than 0.5 gr/l.

3.1.2.13. Enzymatic Determination of Ethanol Content. Boehringer Mannheim/R-Biopharm analysis kit (Catalog number: 10 176 290 035) was used for analysis of ethanol content in samples. The procedure was as follows:

a. 3 ml reaction mixture 3 was pipetted into cuvettes.
b. 0.1ml of sample was pipetted into the cuvettes.
c. The mixture was mixed by pipetting or converting the cuvette .
d. After approximately 3min, absorbance of the mixture was measured at 340 nm using spectrophotometer (A_1).
e. 0.05 ml of solution 3 was added to the mixture, mixed by pipetting.
f. After completion of reaction (approx. 5-10 min), the absorbances were read again (A_2).
g. The same procedure was applied for a standard reaction, 0.1ml of water replaced the sample in the standard reaction.
h. Ethanol contents of the samples were calculated as described in the manual of the kit. The samples were be diluted prior to analysis when the glucose concentration was higher than 0.06 gr/l.

3.1.2.14. Quantification of mRNA expression. The experimental procedures from *Affymetrix GeneChip Expression Analysis Technical Manual* (2000-2003, Affymetrix Inc) were applied in hybridization experiments to quantify mRNA expression levels, the steps followed are as follows (Figure 3.3):

a. Isolation of RNA

b. Isolation of RNA from yeast
c. Precipitation of RNA
d. Quantification of RNA
e. First-strand cDNA synthesis
f. Second-Strand cDNA synthesis
g. Cleanup of Double-Stranded cDNA
h. Synthesis of Biotin-Labeled cRNA
i. Cleanup of Biotin-Labeled cRNA
j. Quantification of the cRNA
k. Checking Unfragmented Samples by Gel Electrophoresis
l. Fragmenting the cRNA for Target Preparation
m. Eukaryotic Target Hybridization
n. Probe Array Wash and Stain
o. Probe Array Scan

3.1.2.15. Mass spectrometry. Metabolome profiles of samples were detected using mass spectrometry. Intracellular samples and extracellular samples were diluted 1:100 and 1:10 in 50 per cent acetonitrile, 50 per cent H_20 , 0.1 per cent formic acid solution prior to analysis with ES-MS. The per cent intensities between 20-1700 m/z of 12 spectra were recorded and binned prior to data analysis. The optimized parameters from Castrillo *et al.* (1997) were used.

3.1.2.16. Buffers and Compositions. The buffers and the solutions are given in Table 3.5.

Table 3.5. Buffers and solutions

Buffer	Composition
LiCl buffer	4M LiCl, 20 mM Tris/HCl, 10 mM EDTA, pH 7.5.
TBE buffer	0.089 M Tris-Borate/0.002 M EDTA, pH 8.3
TAE buffer	0.4 M Tris-Acetate/0.001 M EDTA, pH 8.3
DEPC treated water	0.1 per cent v/v DEPC for at least 1 hour at 37°C, autoclaved
Loading dye	1 per cent bromophenol blue in water
EtBr solution	5.25 mg ethidium bromide /ml water

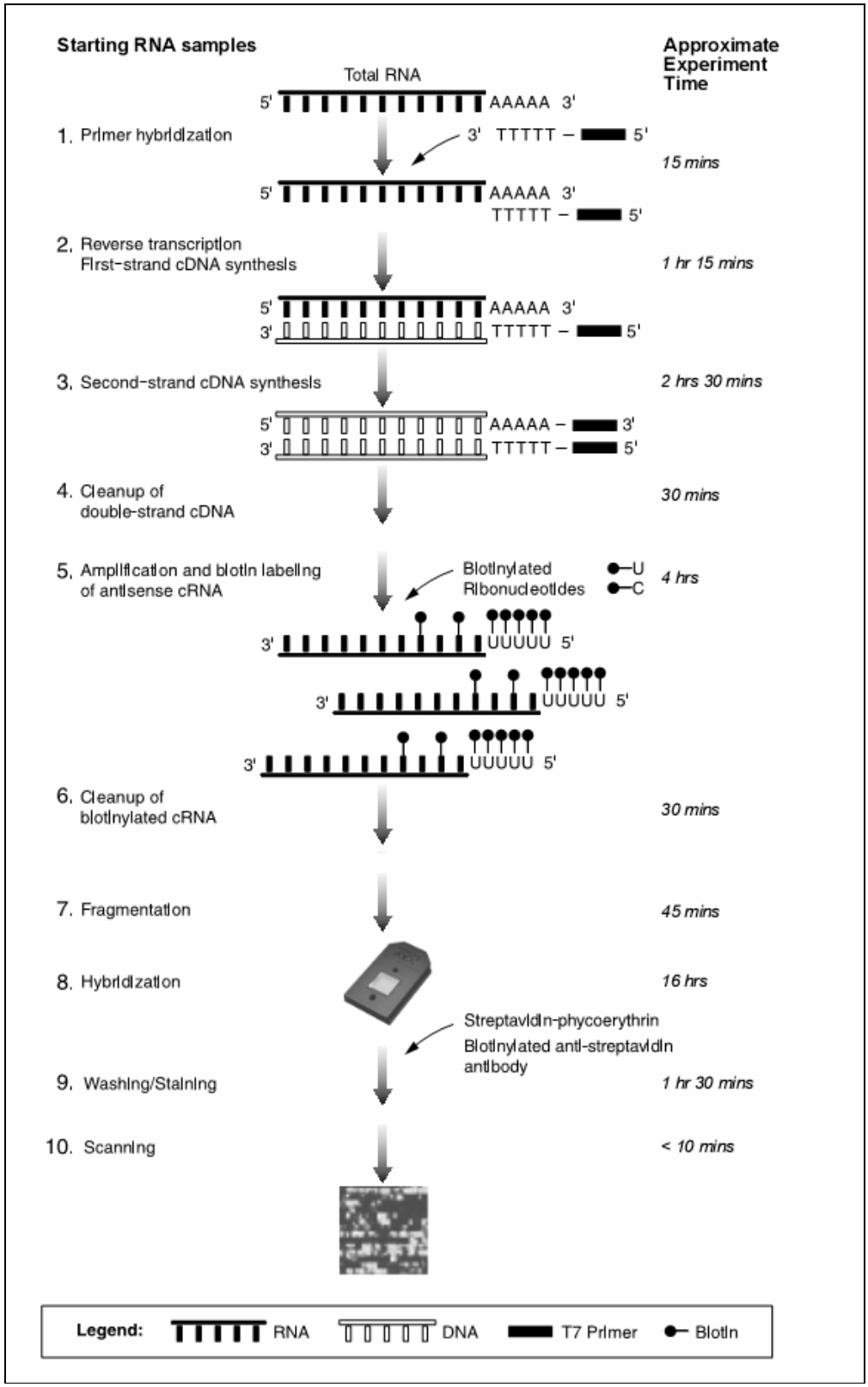

Figure 3.3. Target Labeling for GeneChip Expression Analysis (Affymetrix, 2003).

3.2. Computational Tools and Methods

The softwares used are given in Table 3.6 and the databases referred to for general information and high throughput data are listed in Table 3.7. The online bioinformatics tools used are given in Table 3.8. In the following sections, the mathematical methods used in this study are described in detail.

Table 3.6. Softwares

Software	Supplier
Word	Microsoft Word 2002 (Microsoft Corporation, 1984-2001
Excel	Microsoft Excel 2002 (Microsoft Corporation, 1984-2001)
Matlab	Matlab 6.5.0.180913a Release 13 (Mathworks, 1984-2002)
PLS toolbox	PLS Toolbox Version 3.5 (Eigenvector Research, 1995-2005)
Cytoscape 2.0	Cytoscape 2.0 (Various Institutions - June 2004)
GeneCluster 2.0	GeneCluster Version 2.1.3 Beta Build 20020926 (Whitehead Institute for Biomedical Research, 2002).
Microarray Suite	GeneChip Microarray Suite 5.0.1 (Affymetrix Inc. 2002-2003)
dChip	DNA-Chip Analyzer Version 1.3, (Wong Lab, Harvard School of Public Health and Dana-Farber Cancer Institute, 2000-2004)

Table 3.7. Databases

Database	Referred for:
SGD (Dolinski *et al.*,2005)	General information on genes and proteins, GO annotations of ORFs
CYGD (Güldener *et al.*, 2005)	General information on *S. cereviae* genes and proteins
BIND (Alfarano *et al.*, 2005)	Protein-protein interaction data
GRID (Breitkreutz *et al.*, 2003)	Protein-protein interaction data
KEGG (Kanehisa and Goto, 2002)	Metabolic pathways of *S. cereviae*
STRING (von Mering *et al.*, 2005)	Data on ORFs sharing similarities
YEASTRACT (Yeastract, 2005)	Data on transcription factors and the genes they bind to.

Table 3.8. Online Tools

Tool	Purpose
SGD GO Mapper (Dolinski *et al.*,2005)	Distribution of genes to GO terms
SGD Slim Mapper (Dolinski *et al.*,2005)	Distribution of genes to GO Slim terms
Wu-BLAST2 Gish, W. (1996-2004)	Amino-acid sequence similarity

3.2.6. **Pre-Processing of Transcriptome and Metabolome Data**

Various pre-processing were applied to transcriptome and metabolome data:

a. Normalization by dChip: The software dChip normalizes transcriptome data from all microarrays in an experiment set. The median array is used to normalize all measurements, so that the average measurements from all arrays are closer to each other. This normalization should be done to all arrays that will be analysed together.

b. Log2 transformation: All expression values are transformed to log2 basis. The expression values measured on arrays are not linearly correlated with the actual abundance of mRNA in the cells. Log2 transformation of measured values is applied to obtain a linear correlation between measured values and mRNA abundance.

c. Autoscaling by MATLAB: The transcriptome data is normalized prior to PCA and PLS by MATLAB. The expression values are autoscaled, i.e., average of the measurements is set to 0 and variance of the measurements is set to 1.

d. Autoscaling by GeneCluster: Autoscaling was applied to transcriptome data prior to clustering by GeneCluster.

e. Averaging and binning MS data: The intensity values from 12 spectra were averaged for each sample to obtain the average intensity at each m/z value. The intensities

were then binned (averaged) to obtain a single intensity value for each integer m/z value between 20 and 1700.

3.2.7. 2^k Factorial Experiment Design

Basic 2^3 factorial experiment design was applied to investigate the effects of three factors on *S. cerevisiae* cells. The details of the design, conditions and samples are given in Tables 3.9-10. In the related figures, the symbol '*Δ*' ıs omitted from the sample names.

Table 3.9. 2^3 Factorial Experiment Design

Run	Deletion	Medium	Rate	Sample Name
1	-	-	-	*hoΔ*G1
2	+	-	-	*hap4Δ*G1
3	-	+	-	*hoΔ*N1
4	+	+	-	*hap4Δ*N1
5	-	-	+	*hoΔ*G2
6	+	-	+	*hap4Δ*G2
7	-	+	+	*hoΔ*N2
8	+	+	+	*hap4Δ*N2

Table 3.10. Factors and Experimental Conditions

Factor	Level (-)	Level (+)
Deletion	Homozygous diploid *hoΔ/hoΔ*	Homozygous diploid *hap4Δ/hap4Δ*
Medium	Glucose-limited	Ammonium-limited
Dilution Rate	0.1 h^{-1}	0.2 h^{-1}

3.2.8. Linear Modelling and Analysis of Variance (ANOVA)

The linear model for a factorial experiment with three factors is as follows (Montgomery, 2001):

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl} \qquad (3.1)$$

where i, j, k represent the indices of the levels of the factors; i = 1, 2, ... , a; j = 1, 2, ... , b; k = 1, 2, ... , c; l: index of replicates; l = 1, 2, ..., n; μ : mean of the outputs; y : simulated value of the output variable; τ, β, γ : effects of the factors; ε : random error. The output variable y represents the expression level of an ORF on a $\log_2$ basis.

The linear model in Eq.(3.1) was used to estimate the coefficients to describe the expression level of each ORF. The coefficients obtained in each case are the "effects" of the factors on the expression of the ORF modeled. A positive effect made by a factor indicates that the ORF is up-regulated at Level (+) when compared to Level (-) of that factor. Similarly, a negative effect indicates down-regulation of ORF at Level (+) when compared to Level (-).

Once the model is constructed, analysis of variance (ANOVA) is used to test the hypothesis that the effects are significant or not (equal to zero or not). Analysis of variance partitions the total variability into its parts, *i.e.*, into factors. The ANOVA table for a three-factor model is given in Table 3.11.

Table 3.11. ANOVA table for three-factor model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Fo
Factor A	SS_A	a-1	$MS_A = SS_A/(a-1)$	MS_A/MS_E
Factor B	SS_B	b-1	$MS_B = SS_B/(b-1)$	MS_B/MS_E
Factor C	SS_c	c-1	$MS_C = SS_C/(b-1)$	MS_C/MS_E
Interaction AB	SS_{AB}	(a-1)(b-1)	$MS_{AB} = SS_{AB}/[(a-1)(b-1)]$	MS_{AB}/MS_E
Interaction AC	SS_{AC}	(a-1)(c-1)	$MS_{AC} = SS_{AC}/[(a-1)(c-1)]$	MS_{AC}/MS_E
Interaction BC	SS_{BC}	(b-1)(c-1)	$MS_{BC} = SS_{BC}/[(b-1)(c-1)]$	MS_{BC}/MS_E
Interaction ABC	SS_{ABC}	(a-1)(b-1)(c-1)	$MS_{ABC} = SS_{ABC}/[(a-1)(b-1)(c-1)]$	MS_{ABC}/MS_E
Error	SS_E	abc(n-1)	$MS_E = SS_A/[ab(n-1)]$	
Total	SS_T	abcn-1		

Formulae for sum of squares are as follows:

$$y_{....} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl} \tag{3.2}$$

$$y_{i...} = \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl} \tag{3.3}$$

$$y_{.j..} = \sum_{i=1}^{a} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl} \tag{3.4}$$

$$y_{..k.} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{l=1}^{n} y_{ijkl} \tag{3.5}$$

$$y_{...l} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} y_{ijkl} \tag{3.6}$$

$$y_{ij..} = \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl} \tag{3.7}$$

$$y_{i.k.} = \sum_{j=1}^{b} \sum_{l=1}^{n} y_{ijkl} \tag{3.8}$$

$$y_{.jk.} = \sum_{i=1}^{a} \sum_{l=1}^{n} y_{ijkl} \tag{3.9}$$

$$y_{ijk.} = \sum_{l=1}^{n} y_{ijkl} \tag{3.10}$$

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{abcn} \tag{3.11}$$

$$SS_A = \frac{1}{bcn} \sum_{i=1}^{a} y_{i...}^2 - \frac{y_{....}^2}{abcn} \tag{3.12}$$

$$SS_B = \frac{1}{acn} \sum_{j=1}^{b} y_{.j..}^2 - \frac{y_{....}^2}{abcn} \tag{3.13}$$

$$SS_C = \frac{1}{abn} \sum_{k=1}^{c} y_{..k.}^2 - \frac{y_{....}^2}{abcn} \tag{3.14}$$

$$SS_{AB} = \frac{1}{cn} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij..}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B \tag{3.15}$$

$$SS_{AC} = \frac{1}{bn} \sum_{i=1}^{a} \sum_{k=1}^{c} y_{i.k..}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_C \tag{3.16}$$

$$SS_{BC} = \frac{1}{an}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{.jk..}^2 - \frac{y_{....}^2}{abcn} - SS_B - SS_C \tag{3.17}$$

$$SS_{ABC} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{ijk..}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} \tag{3.18}$$

$$SS_E = SS_T \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} y_{ijk..}^2 - \frac{y_{....}^2}{abcn} \tag{3.19}$$

P-values of the factors were calculated corresponding to F-values obtained via ANOVA in order to indicate the significance of the correlation between the gene expression and the factor.

3.2.9. **Principle Component Analysis (PCA) and Partial Least Squares (PLS)**

In industrial processes, large sets of process data are collected by computerized control and monitoring systems. Multivariate data analysis methods have emerged to compensate for the need for data reduction towards understanding the nature of the process and fault diagnosis. Principal components analysis (PCA) and partial least-squares (projection to latent structures - PLS) are two statistical methods that were proposed for process analysis, monitoring and diagnosis (Geladi and Kowalski, 1986, Kourti and MacGregor, 1995, Wold *et al*., 2001). Later on, these methods were employed as one of the standard tools of chemometrics in analytical chemistry (Hopke, 2003). The theoretical background of PCA (Kourti and MacGregor, 1995) and PLS (Wold, 1987, Geladi and Kowalski, 1986) have been presented and industrial applications reviewed extensively.

PCA was found to be useful in a wide range of applications in bioinformatics, i.e., reducing the dimensions of high-throughput metabolome data (Allen *et al.*, 2003 and Raamsdonk *et al.*, 2001) and identifing the clustering of co-regulated genes (Peterson, 2003) .

PCA projects the original data set to a new space with reduced dimensions. The transformation is made by the loading matrix (p) and the observations are represented by the score matrix (t) in the new space. Columns of p and t matrices correspond to principal components, which lie in the direction of maximum variation that remains in the data after

removal of the variation explained by the previous component. Decomposition of the data matrix X into matrices t, p and the residual matrix e is as follows:

$$X = tp^t + e \tag{3.20}$$

where subscript "*t*" denotes the transpose of the matrix p.

PLS was recently used for the analysis of transcriptome data for classification of samples from human tumors (Nguyen and Rocke, 2002a) and classification of patients for their survival time (Nguyen and Rocke, 2002b). In another study, the genes that are expressed periodically within the cell cycle were determined using PLS (Johansson, 2003).

In PLS methodology, the independent "cause" matrix X and the dependent "response" matrix Y are regressed and modelled simultaneously. The columns of these matrices represent the variables (genes and metabolites in X and Y, respectively) and rows represent the samples. The linear model is:

$$Y = XB + E \tag{3.21}$$

where B is the regression vector and E is the residual matrix.

Projection of the original data set X to a new space with reduced dimensions is made by the loading matrix (p) and the observations are represented by the score matrix (t) in the new space. Decomposition of the data matrix X into the score matrix (t), the loading matrix (p) and the residual matrix (e) is as follows:

$$X = tp^t + e \tag{3.20}$$

where the superscript "*t*" denotes the transpose of the matrix p. Columns of p and t matrices correspond to the latent variables (LVs), which lie in the direction of the maximum variation that remains in the data after removal of the variation explained by the

previous LV. The residual matrix (e) represents the variation that remains unrepresented in t and p matrices. Similarly, the response matrix Y is decomposed as:

$$Y = uq^t + f \tag{3.22}$$

The score vectors (vectors of u) and the loading vectors (vectors of q) correspond to LVs. The residuals are given by the f matrix. A linear inner relation also exists between the matrices t and u, where $\bar{u}$ denotes the matrix of estimated values of u:

$$\bar{u} = bt \tag{3.23}$$

The optimal number of LVs to be included in the model depends on the amount of variation explained by the LVs which are in descending order of the variation they explain. One criterion for the selection of an optimal number of LVs is to set a threshold value for the variation. Then, a sufficient number of LVs are included in the model to represent the threshold variation, and the rest of the variation remains in the residual matrix. Cross-validation is another criterion where the analysis is performed with a subset of the data and the rest of the data set is used to determine the prediction power of the model. Then, the number of LVs that results in minimum prediction error sum of squares (PRESS) is selected.

3.2.10. **GO Mapping**

Gene Ontology mapping tools of Saccharomyces Genome Database (SGD) (Dolinski *et al*., 2005) were used to analyse the biological processes that are significantly common among products encoded by listed ORFs.

GO Slim mapping is composed of a subgroup of GO terms which represent only the major branches of annotations while omitting the details. GO Slim mapping was used whenever an overall picture of distribution of ORFs according to the processes was needed. In order to elucidate the specific processes and functions of ORFs appearing in shorter lists, GO Term Finder mapping was used.

SGD GO Term Finder Tool provides the p-values of the distribution of the ORFs on the ontology terms, however, GO Slim Mapping Tool reports the distribution only. The following equation was used to calculate the p-values for GO Slim Mapping results:

$$p = \sum_{j=x}^{n} \left(\frac{n!}{j!(n-j)!}\right) \phi^{j} (1-\phi)^{(n-j)} \tag{3.24}$$

where $\phi = G/N$

G: ORFs annotated to a specific term.

N: ORFs annotated to all terms.

n: ORFs in the list investigated.

x: ORFs in the list annotated to the specific term.

4. TRANSCRIPTIONAL REGULATION OF CENTRAL CARBON METABOLISM

4.1. Experimental results

In this study, samples from *S. cerevisiae* cells grown in aerated nutrient limited chemostats were collected and investigated. Two sets of data were collected. In the first set, 8 chemostat runs were made to elucidate the effects of three types of perturbations on *S. cerevisiae* cells. The 2^3 factorial experimental design was applied as described in Section 3.2.2, and the perturbations were nutrient limitation (glucose vs ammonium limitation), growth rate (D = 0.1 hr^{-1} vs D = 0.2 hr^{-1}) and gene deletion (*hoΔ/hoΔ* vs *hap4Δ/hap4Δ*). Transcriptome and metabolite (ethanol and glucose) concentrations in samples were analyzed via microarrays and enzymatic methods, respectively. Biomass contents in the chemostat reactors were determined using optical methods (OD600). Transcriptome profiles in this data set were analyzed using linear modeling followed by GO mapping (Section 5). Metabolic data and transcriptome data were integrated using PLS, and the ORFs mediating the effects of these perturbations on metabolic profiles were revealed (Section 5). Novel functional annotations were made for unknown genes significantly responding to perturbations reported both in this work and in literature using correlation of mRNA profiles and GO mapping (Section 6). The transcriptome data were further combined with data from various sources (protein-protein interaction, amino-acid sequence similarity and shared transcription factors) for assigning novel annotations to more than 500 unknown ORFs (Section 7).

In the second data set, only gene deletions were taken into consideration (*hoΔ/hoΔ*, *hap4Δ/hap4Δ, bcs1Δ/bcs1Δ*, *oxa1Δ/oxa1Δ*, *rip1Δ/rip1Δ*, *HAP4*/*hap4Δ*, *RIP1*/*rip1Δ*, *mba1Δ/mba1Δ*, *mig1Δ/mig1Δ*, and *MIG1/mig1Δ*) and 10 chemostat runs were made. The deletion mutants were grown under selected conditions (glucose limitation, D = 0.2 hr^{-1}). Transcriptome and metabolome profiles were determined using microarrays and MS respectively. Ethanol and glucose contents in chemostats were measured via enzymatic methods and biomass content were obtained in terms of OD600. Transcriptome data were

analysed by linear modeling followed by GO mapping, similar to the analysis of transcriptome from the first data set (Section 5). Also, metabolic and metabolome data were integrated with transcriptome using PLS to elucidate the ORFs that mediate the changes in metabolic profiles (Section 5).

Transcriptome data from both sets were combined for investigation of regulation of central carbon metabolism (Section 4).

4.2. Investigation of Transciptional Regulation of Pathways

In this section, the effects of deletion mutations and growth conditions on expression of the genes encoding the central carbon metabolism enzymes of *S. cerevisiae* are investigated. Homozygous deletion mutants of *HAP4* and HO were grown under glucose and ammonia limitation and at dilution rates 0.1 hr^{-1} and 0.2hr^{-1}. *HAP4*, *RIP1* and *MIG1* deletions were reported to have haploid insufficiency (Delneri, 2002), thus both homozygous and heterozygous deletion mutants of *HAP4*, *RIP1* and *MIG1* were grown under selected conditions (glucose limitation at dilution rate 0.2hr^{-1}). Only homozygous deletion mutants of *BCS1*, *OXA1* and *MBA1* were grown under selected conditions. Homozygous HO deletion mutants were used as reference strains, glucose limitation was used as the reference growth medium and dilution rate 0.1hr^{-1} was used as the reference growth rate. In Section 2.8 a literature survey on functions of the deleted genes is presented.

240 genes encoding the central metabolism enzymes were selected from the complete *S. cerevisiae* genome. The effect of each deletion or condition on the gene expression with respect to reference deletion/condition was calculated using linear modeling.

A general picture of the central metabolism is given in Figure 4.1 with the expression profiles of the selected genes. The statistical significance of the effects was not considered, hence only the magnitudes of the effects are illustrated. Each part of the metabolism is then investigated in detail, and comparison of the present results with the results from literature is also provided.

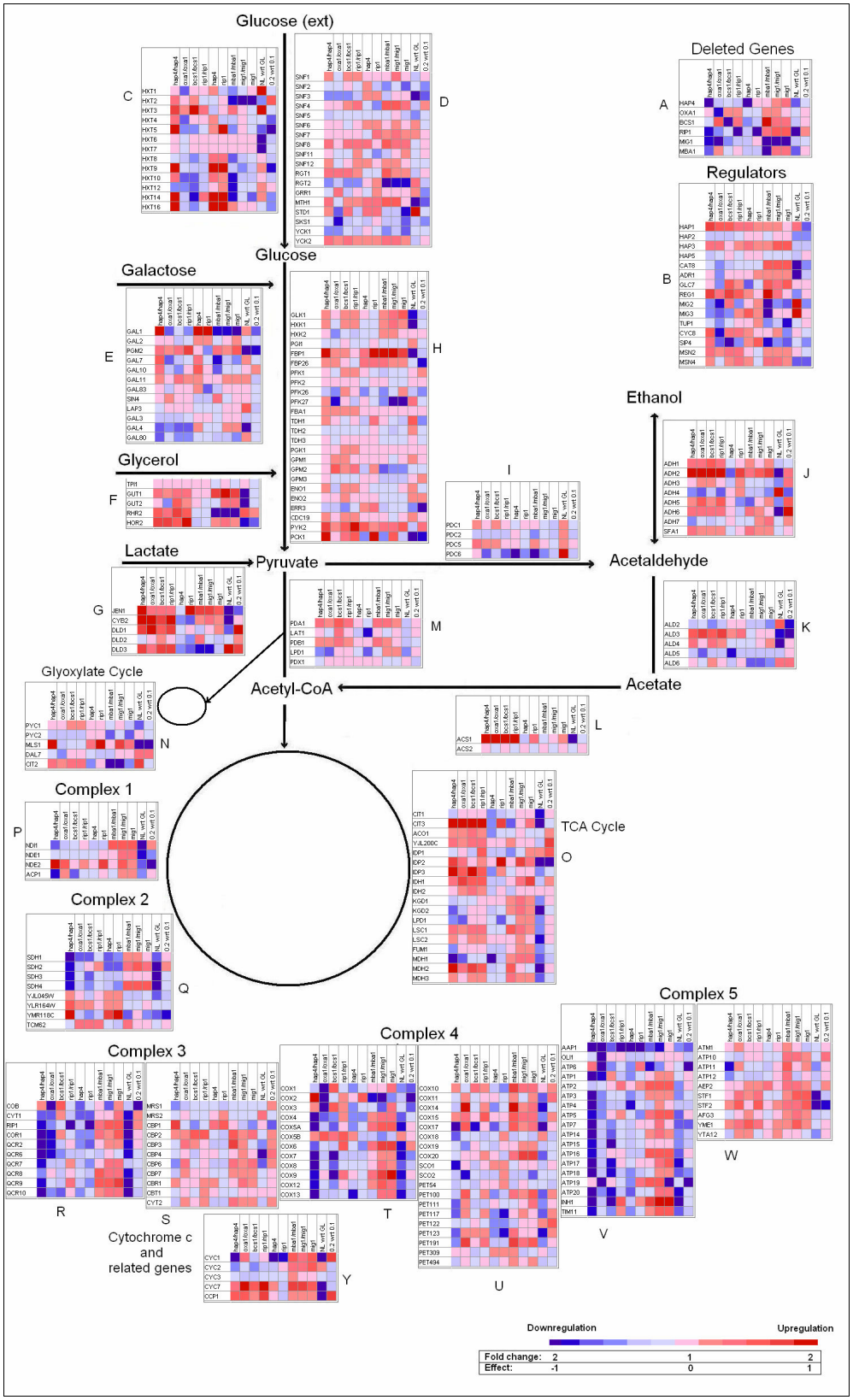

Figure 4.1. Expression profiles of selected genes.

The deletions and conditions are given on the columns of the color-coded expression boxes in each step of central metabolism in Figure 4.1. The genes involved in the corresponding step of the metabolism are listed in the rows. Blue color indicates that the gene given in the row is downregulated at the condition or deletion mutant specified at the top of the column. Similarly, red color implies upregulation. Darker colors indicate strong effects while lighter colors indicate weaker effects. The color scale can be considered as a representation of fold changes as shown in Eqs.4.1-4.

$$\text{Fold change} = \frac{\text{Expression(A)}}{\text{Expression(B)}} \tag{4.1}$$

$$\text{Effect} = \log_2\left[Expression(A)\right] - \log_2\left[Expression(B)\right] \tag{4.2}$$

$$\text{Effect} = \log_2\left[\frac{Expression(A)}{Expression(B)}\right] \tag{4.3}$$

where Expression(A) represents the average expression under condition A and Expression(B) represents the average expression under condition B. Thus,

$$\text{Effect} = \log_2\left[Foldchange\right] \tag{4.4}$$

In the color scale, effects with magnitude larger than |1| are represented by darkest colors (red or blue) while intermediate effects between [1, -1] are represented by intermediate colors. Similarly, the darkest colors (red or blue) on the scale represent upregulation or downregulation more than 2-fold while fold changes between 1 and 2 are represented with intermediate colors.

4.2.1. Phenotype of Deletion Mutants and Expression Profiles of Deleted Genes

Homozygous deletion mutants *ho/hoΔ* (reference strain), *hap4Δ/hap4Δ*, *oxa1Δ/oxa1Δ*, *bcs1Δ/bcs1Δ*, *rip1Δ/rip1Δ*, *mig1Δ/mig1Δ*, and *mba1Δ/mba1Δ* and heterozygous deletion mutants *HAP4*/*hap4Δ*, *RIP1*/*rip1Δ* and *MIG1*/*mig1Δ* were

investigated in this study. The heterozygous deletion mutants *HAP4*/*hap4Δ*, *RIP1*/*rip1Δ* and *MIG1*/*mig1Δ* were selected because of haplo-insufficiency observed in these strains (Delneri, 2004). In a diploid cell, heterozygous deletion (deletion of a gene from only one allele) would be expected to cause no phenotype; however, insufficient concentration of the gene product may cause inhibition of growth and cell division at normal rate, which phenomenon is known as haplo-insufficiency (Oliver, 1999).

Homozygous deletions *hap4Δ/hap4Δ*, *bcs1Δ/bcs1Δ*, *oxa1Δ/oxa1Δ* and *rip1Δ/rip1Δ* cause complete respiratory deficiency in *S. cerevisiae* cells (verified by inhibited growth on ethanol, Appendix A). Heterozygous deletion mutants *HAP4*/*hap4Δ* and *RIP1*/*rip1Δ* and homozygous deletion strain *mba1Δ/mba1Δ* have partial respiratory deficiency. Homozygous and heterozygous deletions *mig1Δ/mig1Δ* and *MIG1/mig1Δ* cause no growth deficiency under the conditions investigated. The transcriptional response of central carbon metabolism to heterozygous deletion of a gene is expected to be in the same direction as the homozygous deletion of the gene. Homozygous and heterozygous deletion mutants of *HAP4* and *MIG1* exhibit this expected behavior in this study, however, exceptionally a group of transport genes response in mutant *RIP1*/*rip1Δ* but not in *rip1Δ/rip1Δ* (see the discussion below).

A transcriptional regulation cascade is proposed for the genes whose deletion mutants are used in this study. The present expression data given in color code in Figure 4.1A imply the regulation cascade given in Figure 4.2. Red arrows indicate activation and blue arrows indicate repression. Expression of each gene is very strongly downregulated in its own deletion mutant, as expected; this behavior is not illustrated in Figure 4.2. Expression of *MIG1*, *MBA1* and *RIP1* were downregulated in strains *hap4/hap4Δ* and *HAP4*/*hap4Δ*; thus Hap4p may be an activator of *MIG1*, *MBA1* and *RIP1* expression. Activation of *RIP1* is expected since Hap4p activates the expression of respiration genes and *RIP1* has Hap4p binding site. Activation of *MIG1* and *MBA1* expression may be “feedback control”, because the presence of *MIG1* and *MBA1* genes in the genome is found to be repressing respiration as will be discussed. Probably, *MIG1* and *MBA1* expressions are activated by Hap4*p* to repress the overexpression of respiration genes. *MBA1* has Hap4p binding site but *MIG1* does not.

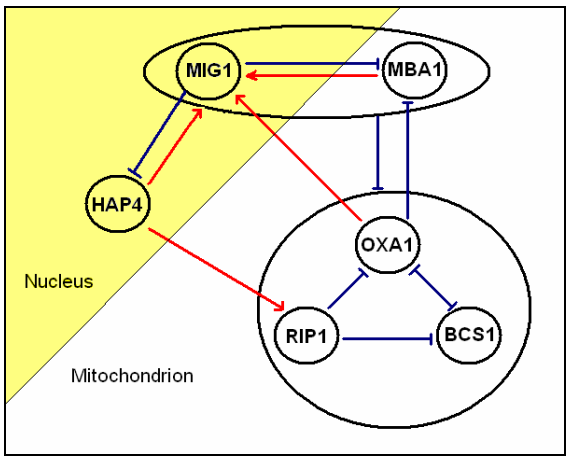

Figure 4.2. Proposed regulation cascade (red: activation, blue: repression)

MIG1 expression was downregulated in *mba1Δ/mba1Δ* deletion mutant and *MBA1* expression was upregulated in *mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletion mutants. Thus, presence of Mba1p may activate *MIG1* expression and existence of Mig1p may repress *MBA1* expression. The deletions *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* affect other genes in the same direction. Consequently, activation of *MIG1* by Mba1p and repression of *MBA1* by Mig1p may be a feedback control loop to control the level of regulation exerted by Mig1p on the other genes.

Retrograde regulation is defined as the nuclear response to reduction of mitochondrial function. Absence of Oxa1p was reported by Lemaire *et al.* (2000) to be inducing levels of two proteins participating in retrograde signaling from mitochondria to nucleus (Pdr3p and Pdr5p), and *OXA1* loci were found to be a negative regulator of *PDR3* and *PDR5.* The results of this study also verify that the expressions of *PDR3* and *PDR5* are induced in deletion mutant *oxa1Δ/oxa1Δ* (results not shown).

MIG1 expression was downregulated in *oxa1Δ/oxa1Δ* deletion mutant. Main effect of deletion *oxa1Δ/oxa1Δ* was found to be the downregulation of mitochondrially encoded respiration genes in this work, and Oxa1p was reported to be involved in membrane insertion machinery (Stuart, 2002). These findings are irrelevant to the regulation of *MIG1* expression. Moreover, if *MIG1* expression was positively regulated by Oxa1p, expression profile of deletion mutant of *oxa1Δ/oxa1Δ* would be expected to be similar to those of

mutants *mig1Δ/mig1Δ* and *MIG1*/*mig1Δ*. It is very probable that downregulation of *MIG1* expression in *oxa1Δ/oxa1Δ* mutant is a false positive in terms of transcriptional regulation. Similarly upregulation of *MBA1* expression in *oxa1Δ/oxa1Δ* mutant may be a false positive.

In the strains with *mig1Δ/mig1Δ* and *MIG1/mig1Δ* deletions, expressions of *HAP4*, *BCS1*, *OXA1*, *MBA1* and *RIP1* were upregulated, thus Mig1p may be a repressor of these genes. Only *HAP4* is known to be regulated by Mig1p. Repression of *MBA1* by Mig1p may be a feedback control loop as mentioned above. Repression of *HAP4*, *BCS1*, *OXA1* and *RIP1* is expected since these genes are known to be respiration related. Similarly, Mba1p may be a repressor of *BCS1*, *OXA1* and *RIP1*. It may also repress *HAP4*, but the effect is weak. Probably, the repression is not directly done by Mba1p, but done through activation of *MIG1* expression by *MBA1*. Otherwise, high level of *MBA1* in *mig1Δ/mig1Δ* and *MIG1/mig1Δ* deletions would be adequate to cause downregulation of these genes. However, these genes are upregulated in *MIG1* deletion, thus *MIG1* should be needed to apply the repression.

In the mutant *rip1Δ/rip1Δ*, expressions of *BCS1* and *OXA1* are upregulated. In the mutants *bcs1Δ/bcs1Δ* and *oxa1Δ/oxa1Δ*, *OXA1* and *BCS1* expressions are upregulated respectively. *OXA1* and *BCS1* expressions are upregulated by almost all deletions causing respiratory deficiency; thus, their upregulation may be a result of respiratory deficiency but not a result of absence of the deleted gene.

The last two columns of Figure 4.1A give the changes in gene expression as response to medium and growth rate change. “NL wrt GL” column shows the response to ammonium limitation as compared to glucose limitation. The glucose concentration is high in ammonium limited media, thus the main effect of ammonium limitation on central carbon metabolism is thought to be glucose-dependent.

HAP4, *RIP1* and *MBA1* expressions are downregulated when the glucose level is high when compared to glucose limitation. This response may indicate the glucose repression exerted on the expression of these genes by the ammonium limited media which

has high glucose concentration. As may be seen in the last column of Figure 4.1A, *RIP1* expression is upregulated at higher growth rate (0.2 hr^{-1}) when compared to low growth rate (0.1 hr^{-1}). The responses of other genes to medium and growth rate changes are insignificant.

4.2.2. Various Regulators of Central Metabolism

The genes encoding proteins of HAP complex and genes that were reported to be acting on carbon catabolite repression (Gancedo, 1998) are given in Figure 4.1B.

HAP1 and *HAP3* are upregulated by all deletions, while *HAP2* and *HAP5* are not strongly affected by any deletion. Expressions of *MSN2, MSN4* and *REG1* are upregulated by all deletions.

MIG2 and *CYC8* expressions may be activated by Oxa1p while *MIG2*, *MIG3* and *SIP4* may be activated by *MBA1*. *SIP4* may also be activated by *BCS1*.

GLC7 is upregulated by *bcs1Δ/bcs1Δ*, *rip1Δ/rip1Δ, mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions and downregulated by *oxa1Δ/oxa1Δ* deletion.

The deletions *hap4Δ/hap4Δ*, *oxa1Δ/oxa1Δ* and *mba1Δ/mba1Δ* downregulate *MIG2* and *MIG3* expression. They are not affected strongly in strains *mig1/mig1Δ* and *MIG1*/*mig1Δ* but are strongly upregulated by high glucose content. Mig2p and Mig3p have aminoacid sequence similarity to Mig1p, but they were reported to have limited or no contribution to glucose repression (Gancedo, 1998). The results of the present study demonstrate that *MIG2* and *MIG3* may be transcriptionally regulated by Mba1p and may be involved in glucose repression.

MIG1 expression is not affected under ammonium limitation while *CAT* and *ADR1* expressions are strongly downregulated.

Growth rate change does not affect these regulators significantly.

4.2.3. Hexose Transporters

Most of the hexose transporters are upregulated by deletions *hap4Δ/hap4Δ* (Figure 4.1C), even heterozygous deletion *HAP4*/*hap4Δ* upregulates their expression. Also many of them are upregulated at heterozygous *RIP1/rip1Δ* deletion but are not downregulated by homozygous *rip1Δ*/*rip1Δ* deletion. *HXT5*, *HXT9*, *HXT14* and *HXT16* are strongly upregulated in *hap4Δ*/*hap4Δ, HAP4*/*hap4Δ* and *RIP1*/*rip1Δ*. This behavior is interesting because these genes are not affected by homozygous *rip1Δ*/*rip1Δ* deletion. *HXT3* is the only hexose transporter that is upregulated at all respiratory deficient conditions. *HXT2* is downregulated by deletions of *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1/mig1Δ*, while *HXT10*, *12* and *14* are downregulated by deletion of *mba1Δ/mba1Δ* only.

HXT1 and *HXT3* are strongly upregulated in the presence of glucose (ammonium limitation case) while *HXT6* and *HXT7* are strongly downregulated. This behaviour supports the hypothesis that *HXT1* and *HXT3* are low-affinity glucose transporters that are active when glucose concentration is high, and *HXT6* and *HXT7* are high-affinity glucose transporters that are active when glucose concentration is low (Özcan and Johnston, 1999).

Increasing the growth rate upregulates *HXT2* and downregulates *HXT5* expression strongly.

4.2.4. Glucose Signaling Pathway

The effect of the deletions on genes involved in glucose signaling pathway is not too strong (Figure 4.1D). Most of these genes are affected positively from the deletions, which may indicate that these genes are slightly downregulated by the presence of the deleted genes. An exception is *RGT2*, which is downregulated by *mba1Δ/mba1Δ, mig1/mig1Δ* and *MIG1*/*mig1Δ* deletions, *RGT2* may be activated by products of these genes. Another exception is *SKS1* which may be activated by Oxa1p. High glucose level in ammonium limited medium downregulate *SNF3* and *MTH1* expression, while the expression of *STD1* is strongly upregulated. Change in growth rate does not have a significant effective on these genes.

4.2.5. Galactose Pathway

GAL1 exhibits similar behavior to some of the hexose transporters summarized above; it is strongly upregulated by *hap4Δ/hap4Δ, HAP4/hap4Δ* and heterozygous *RIP1/rip1Δ* deletions (Figure 4.1E). It is downregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions. *PGM2* (*GAL5*) expression is upregulated in deletion mutants that are used in this study, except for heterozygous *RIP*1/*rip1Δ* deletion. *GAL11* is also upregulated by all deletions. Other genes are not strongly affected from the deletions.

PGM2 and *GAL4* are downregulated when glucose level is high. Other genes are either upregulated at ammonium limitation or do not respond. *LAP3* (*GAL6*) is upregulated more significanty then others. Increase in the growth rate downregulates *GAL1* and *PGM2* while upregulating *GAL10*.

4.2.6. Glycerol metabolism

All genes on glycerol metabolism pathway (Fig 4.1F) are upregulated in respiratory deficient mutants (except for *GUT2* which is not upregulated by *oxa1Δ/oxa1Δ* deletion). *GUT1* may be strongly downregulated by *MBA1* and *MIG1* while *RHR2* (*GPP1*) is strongly activated. *RHR2* is upregulated when glucose level and grow rate are high, while *GUT1*, *GUT2* and *HOR2* (*GPP2*) are either downregulated or unchanged by these conditions. Rhr2p and Hor2p catalyze production of glycerol, contrary to Gut1p and Gut2p which catalyze the reactions in metabolism of glycerol.

Deletion of *tpi1Δ*, *gut1Δ* and *gut2Δ* were reported to cause growth deficiency on non-fermentable carbon sources, which indicate respiratory deficiency; on the contrary, *rhr2Δ* (*gpp1Δ*) deletion mutant was reported to exhibit growth deficiency on fermentable carbon sources (Steinmetz *et al.*, 2002). *RHR2* was also reported to be upregulated at anaerobic conditions while *HOR2* expression was not affected (Pahlman *et al*., 2001). Thus, Rhr2p, an isoform of DL-glycerol-3-phosphatase Hor2p, enables metabolism of glycerol when glucose repression or anaerobic conditions apply. Under these conditions, fermentation

dominates the metabolism and induction of Rhr2p indicates that it is needed for metabolism of glycerol when these conditions apply. Consequently, its deletion is expected to cause growth defect on fermentable substrates. The response *TPI1* was insignificant in present gene deletions and conditional changes.

4.2.7. Lactate Metabolism

All genes from lactate metabolism except *DLD2* are upregulated in respiratory deficient mutants (Figure 4.1G). *JEN1*, which is a lactate transporter is also affected from heterozygous *RIP1*/*rip1Δ* deletion but not from heterozygous *HAP4*/*hap4Δ* deletion, similar to previously discussed transporters. This gene is also upregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1/mig1Δ* deletions. *CYB2* is also upregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions while *DLD3* is downregulated by *mba1Δ/mba1Δ* and *mig1Δ/mig1Δ* deletions.

DLD1-2-3, D-lactate dehydrogenases converting lactate to pyruvate, are upregulated at higher growth rate and *DLD3* is also upregulated by high glucose level in ammonium limited conditions. Dkd1p and Dld2p are mitochondrial and Dld3p is cytosolic.

CYB2 expression is downregulated by glucose and anaerobic conditions (Lodi, 1991) and Cyb2p is located in mitochondria. *JEN1* is a lactate transporter which is required for uptake of lactate and pyruvate; its expression is derepressed by transcriptional activator Cat8p under nonfermentative growth conditions and downregulated in the presence of glucose, fructose, and mannose (Haurie, 2001). *JEN1* and *CYB2* are downregulated when glucose level is high in the present data, and this result is in agreement with Lodi *et al.,* (1991) and Haurie *et al.,* (2001)

4.2.8. Glycolysis

The genes encoding the glycolysis enzymes are generally downregulated in respiratory deficient mutants, though this effect is weak for most of them (Figure 4.1H). Expressions of *GLK1, HXK1, HXK2, FBP1, FBP26* and *PYK2* are upregulated by

mba1Δ/mba1Δ, mig1Δ/mig1Δ and *MIG1*/*mig1Δ* deletions. *PFK27* expression is downregulated by *oxa1Δ/oxa1Δ, mig1Δ/mig1Δ* and *MIG*1/*mig1Δ* deletions.

Deletion mutants of *hxk2Δ, pfk1Δ, pfk2Δ* and *tdh3Δ* were reported to be growth deficient on fermentable carbon sources, while *fbp1Δ* deletion was reported to be growth deficient on non-fermentable carbon sources (Steinmetz *et al.,* 2002).

Expressions of *GLK1, HXK1* and *PCK1* are downregulated at high glucose levels under ammonium limited conditions, while other genes are either unchanged or upregulated when glucose level is high. Thus, glucose repression is effective only on *GLK1, HXK1* and *PCK1* expressions in the present results. *FBP26, ERR3* and *PCK1* expressions are downregulated at the higher growth rate, while other genes are either unchanged or upregulated.

4.2.9. From Pyruvate to Acetaldehyde

These genes are not strongly affected from the deletions and growth rate change. *PDC1* and *PDC5* expressions are upregulated in respiratory deficient mutants (Figure 4.1I). *PDC6* is downregulated by heterozygous *HAP4*/*hap4Δ* and homozygous *mba1Δ/mba1Δ* deletions.

All genes encoding the proteins catalysing production of acetaldehyde from pyruvate are upregulated when glucose level is high, which is an expected response. Production of acetaldehyde from pyruvate may indicate that fermentation dominates the metabolism.

4.2.10. From Acetaldehyde to Ethanol

The genes encoding the alcohol dehydrogenases that catalyse the production of ethanol from acetaldehyde are generally upregulated by the deletions and higher growth rate (Figure 4.1J). *ADH2* expression was stongly upregulated in respiratory deficient mutants; *ADH1*, *ADH5*, *ADH5* and *STF1* expression are also upregulated in respiratory deficient mutants. *ADH1* deletion was reported to be growth deficient on fermentable carbon sources (Steinmetz *et al.,* 2002).

ADH2 and *ADH6* were strongly downregulated when glucose level was high while *ADH4* was strongly upregulated, which may indicate that *ADH4* is the active isoform in cells under glucose repression while *ADH2* and *ADH6* are inactive.

4.2.11. From Acetaldehyde to Acetate

ALD4 and *ALD5* are mitochondrial aldehyde dehydrogenases, while others are cytosolic (Figure 4.1K). *ALD3* and *ALD4* are upregulated in respiratory deficient mutants and downregulated when glucose level is high. The other genes are not strongly affected by the deletions. *ALD2* is upregulated when glucose level is high and downregulated by high growth rate.

ALD6 deletion was reported to be growth deficient on fermentable carbon sources (Steinmetz *et al.,* 2002). *ALD2* and *ALD3* were defined as stress inducable and glucose downregulated (Navarro-Avino *et al.*, 1999).

4.2.12. From Acetate to AcCoA

These genes act on the pathway from acetate to AcCoA (Figure 4.1L). *ACS1* is very strongly upregulated in respiratory deficient mutants and downregulated when glucose level is high. *ACS1* deletion was reported to be growth deficient on non-fermentable carbon sources (Steinmetz *et al.,* 2002).

4.2.13. From Pyruvate to AcCoA

The components of the enzyme complex that catalyses the reaction from pyruvate to AcCoA is not strongly affected from the deletions (Figure 4.1M). *PDA1* and *PDB1* are upregulated by almost all deletions, including upregulation in respiratory deficient mutants. These genes were not significantly affected from ammonium limitation or growth rate change. Deletion mutants of these genes were reported to be growth deficient on non-fermentable carbon sources (Steinmetz *et al.,* 2002).

4.2.14. Glyoxylate Cycle

PYC1, *CIT2* and *PCK1* are upregulated in respiratory deficient mutants, although the effect is not too strong for *PYC1*. *CIT2*, encoding the enzyme for production of citrate from oxaloacetic acid, is upregulated when glucose level is high and downregulated by *mba1Δ/mba1Δ* and *mig1Δ/mig1Δ* deletions (Figure 4.1N).

Both *MLS1* and *PCK1* exhibit the behavior of transporters discussed above. In addition, they are downregulated when the glucose level and growth rate are high.

Deletion mutants *pck1Δ* and *mls1Δ* were reported to be growth deficient on non-fermentable carbon sources (Steinmetz *et al.,* 2002). *PYC1* and *PYC2* are both cytosolic pyruvate carboxylases producing oxaloacetic acid from pyruvate, however, their regulation differs from each other (Brewster *et al.*, 1994).

4.2.15. TCA Cycle

Most of the genes encoding enzymes of TCA cycle are upregulated in respiratory deficient mutants, the only exception is *MDH1*, whose expression is downregulated by *hap4Δ/hap4Δ*, *HAP4*/*hapΔ* and *oxa1Δ*/*oxa1Δ* deletions (Figure 4.1O).

The TCA cycle genes are also upregulated by *MIG1*/*mig1Δ* and *mig1/mig1Δ* deletions and downregulated at high level of glucose except for *IDP1* and *IDH2*, which is upregulated at high level of glucose. *IDP2* is downregulated at high growth rate, all the other genes are either unchanged or upregulated by the growth rate.

Deletion mutants of most of these genes were reported to be growth deficient on non-fermentable carbon sources (Steinmetz *et al.,* 2002).

4.2.16. Respiratory Chain Complex I

Nde1p and Nde2p are cytosolic isoenzymes of Ndi1p which replaces the Complex I in *S. cerevisiae* (Figure 4.1P). *NDE2* is upregulated by all deletions, while the others are

upregulated by *mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions and downregulated by high levels of glucose. *NDI1* is also upregulated by *mba1Δ/mba1Δ* deletion. *NDI1* and *ACP1* may be activated by *Hap4p*.

Deletion mutant ndi1*Δ* was reported to be growth deficient on non-fermentable carbon sources (Steinmetz *et al.,* 2002).

4.2.17. Respiratory Chain Complex II

SDH1-4, the genes encoding the components of Complex II (succinate dehydrogenase), are upregulated by *mba1Δ/mba1Δ*, *mig1Δ/mig1Δ* and *MIG1/mig1Δ* deletions and downregulated by *hap4Δ/hap4Δ* deletion and high glucose level (Figure 4.1Q).

The genes other than *SDH1-4* are reported to be related to Complex II activity by KEGG database, their expression behavior is different from the well-known components *SDH1-4*. *TCM62* gene encodes a chaperone necessary for the assembly of the mitochondrial succinate dehydrogenase (Dibrov et al. 1998). Products of *YMR118c*, *YJL045w* and *YLR164w* have amino-acid sequence similarity to Sdh1-4p.

YMR118C was reported to have specifically higher expression in carbon limited chemostat cultures versus carbon excess (Boer *et al.*, 2003), indeed, it was downregulated in high glucose medium in present study. *YMR118C* behaves similar to transporters strongly upregulated by *hap4Δ*/*hap4Δ, HAP4*/*hap4Δ* and heterozygous *RIP1*/*rip1Δ* deletions. Also *YJL045w* and *YLR164w* expression also exhibit the similar behavior but inductions are not strong.

The four deletion mutants sdh1-4Δ was reported to be growth deficient on non-fermentable carbon sources (Steinmetz *et al.,* 2002). Deletion mutants of all protein complexes in respiratory chain was reported to be respiratory deficient by Steinmetz *et al.* (2002).

4.2.18. Respiratory Chain Complex III

All Complex III genes may be activated by Hap4p, except for *COB*, the mitochondrially encoded cytochrome b (Figure 4.1R). *COB* is upregulated by *hap4Δ/hap4Δ* and *bcs1Δ/bcs1Δ* deletions and downregulated by *oxa1Δ/oxa1Δ* deletion and high growth rate. *oxa1Δ/oxa1Δ* deletion also downregulates some of the nuclearly encoded components.

Most Complex III compoenents are upregulated in *mba1Δ/mb1aΔ, mig1Δ/mig1Δ* and *mig1Δ* strains and downregulated when glucose level is high. *COR1* and *CYT1* are downregulated by *bcs1Δ/bcs1Δ* deletion and *QCR7* and *QCR9* may be downregulated by *rip1Δ/rip1Δ*.

4.2.19. Regulation of Respiratory Chain Complex III

The genes that are proposed to be involved in regulation of Complex III are given in Figure 4.1S. They are generally upregulated by deletions and high growth rate.

CBP1 behaves similar to transporter genes, it is upregulated by *hap4Δ/hap4Δ, HAP4/hap4Δ* and heterozygous *RIP1*/*rip1Δ* deletions.

CBP2 is upregulated in respiratory deficient mutants and *mba1Δ/mba Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions, *CBR1* behaves similar to *CBP2*. *CBP4* is downregulated when glucose level is high.

4.2.20. Cytochrome c Related Genes

Cyc1p and Cyc7p are isoforms of cytochrome c. Cyc3p is holocytochrome-c synthase. Ccp1p is cytochrome-c peroxidase precursor. Cyc2p is cytochrome-c mitochondrial import factor. These genes are all upregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions (Figure 4.1Y).

CYC1 and *CCP1* are downregulated when glucose level is high and upregulated by higher growth rate. *CYC7* and *CCP1* are strongly upregulated in respiratory deficient strains, *CYC7* is downregulated when glucose level is high.

CYC1 exhibits the opposite behavior of some transporters discussed above. It is upregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions and downregulated when glucose level is high.

4.2.21. Respiratory Chain Complex IV

These genes encode the components of Complex IV.1. *COX1*, *COX2* and *COX*3 are mitochondirally encoded, and they are all downregulated by *oxa1Δ/oxa1Δ* (Figure 4.1T).

COX2 behave similar to some transporters which are upregulated by *hap4Δ/hap4Δ, HAP4/hap4Δ* and heterozygous *rip1Δ* deletions. In addition, *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions downregulates expression of *COX2*.

The other components, which are nuclearly encoded, are downregulated by *hap4Δ/hap4Δ* and high glucose level; upregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions except for *COX5B*.

COX5Bp is the aneorobic isoform of COX5Ap and *COX5B* expression is upregulated in respiratory deficient mutants.

COX1 and *COX3* are downregulated while *COX7* and *COX9* upregulated by *bcs1Δ/bcs1Δ* deletion.

4.2.22. Regulation of Respiratory Chain Complex IV

The genes that are proposed to have a regulatory action on Complex IV are given in Figure 4.1U.

COX10 and *COX15* are upregulated in *hap4Δ/hap4Δ* deletion in addition to being upregulated in other respiratory deficient strains. They are also upregulated by *mba1Δ/mba Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions.

COX14 is very strongly upregulated by *mba1Δ/mba1Δ* and *oxa1Δ/oxa1Δ* deletions. *SCO2* is strongly downregulated by *oxa1Δ/oxa1Δ* though it is not mitochondrially encoded.

COX17 is downregulated by *hap4Δ/hap4Δ* and *bcs1Δ/bcs1Δ* deletions and high glucuse level. *COX18* and *SCO1* are downregulated by *hap4Δ/hap4Δ* deletion.

COX10-15, PET100, PET123 and *PET121* are upregulated by *mba1Δ/mba1Δ, rip1Δ/rip1Δ* and *oxa1Δ/oxa1Δ* deletions similar to *COX14*, although the effect is weaker.

mig1Δ/mig1Δ, *MIG1*/*mig1Δ, mba1Δ/mba1Δ, rip1Δ/rip1Δ* and *oxa1Δ/oxa1Δ* deletions generally induce expression of these genes.

4.2.23. Respiratory Chain Complex V

Mitochondrially encoded *AAP1* is downregulated by almost all deletions (Figure 4.1V). Other mitochondrially encoded components, *ATP6* and *OLI1* are downregulated by *oxa1/oxa1Δ* deletion. *ATP6* is also downregulated by *mba1Δ/mba1Δ* deletion.

The other components, which are nuclearly encoded, are downregulated by *hap4Δ/hap4Δ*. Some are also downregulated by *bcs1Δ/bcs1Δ* and some are upregulated by *rip1Δ/rip1Δ*. Most of nuclear genes are upregulated by *mba1Δ/mba1Δ*, *mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions.

High glucose level generally downregulate these genes, *ATP6* and *ATP19* are also downregulated by high growth rate. *ATP19* is also downregulated by *oxa1Δ/oxa1Δ* deletion though it is a nuclearly encoded gene.

4.2.24. Regulation of Respiratory Chain Complex V

The genes that are proposed to have role in regulation of Complex V are given in Figure 4.1W. All of them are generally upregulated by the deletions except for *ATP11*, which is dowregulated by *hap4Δ/hap4Δ* deletion. *STF1* and *STF2* were downregulated when glucose level is high which may indicate that they are subject to glucose repression.

4.3. Clustering of Genes Involved in Central Carbon Metabolism

As discussed above, the response of the genes to deletions and conditions vary even for the genes involved in the same pathway or same protein complex. Hierarchical clustering is applied to illustrate the genes with similar expression profiles as neighbors. The hierarchical tree is given on the left and the expression profiles are given on the right side in Figure 4.3. Clustering by self organizing maps (SOM) is applied to find the expression motifs common among the genes from central metabolism of yeast. In Figure 4.4, the average expression of the genes in each cluster is given with the confidence intervals. Almost each cluster has a distinct expression motif (Table 4.1), biological significance of each motif is discussed below. Distribution of the genes in each cluster to metabolic pathways are given in Table 4.2.

Table 4.1. Expression motifs in central carbon metabolism

Cluster	Expression Motif
0	Upregulation when glucose level is high
1	Downregulation by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions.
2	Upregulation by *hap4Δ/hap4Δ, HAP4/hap4Δ* and *RIP1*/*rip1Δ* deletions.
3	Downregulation by *oxa1Δ/oxa1Δ* deletion
4	Upregulation in respiratory deficient mutants and high growth rate
5	Upregulation in respiratory deficient mutants, downregulation when glucose level and growth rate are high.
6	Downregulation by *hap4Δ/hap4Δ* deletion and high glucose level.
7	Downregulation by *hap4Δ/hap4Δ* deletion and high glucose level.
8	Downregulation when glucose level is high

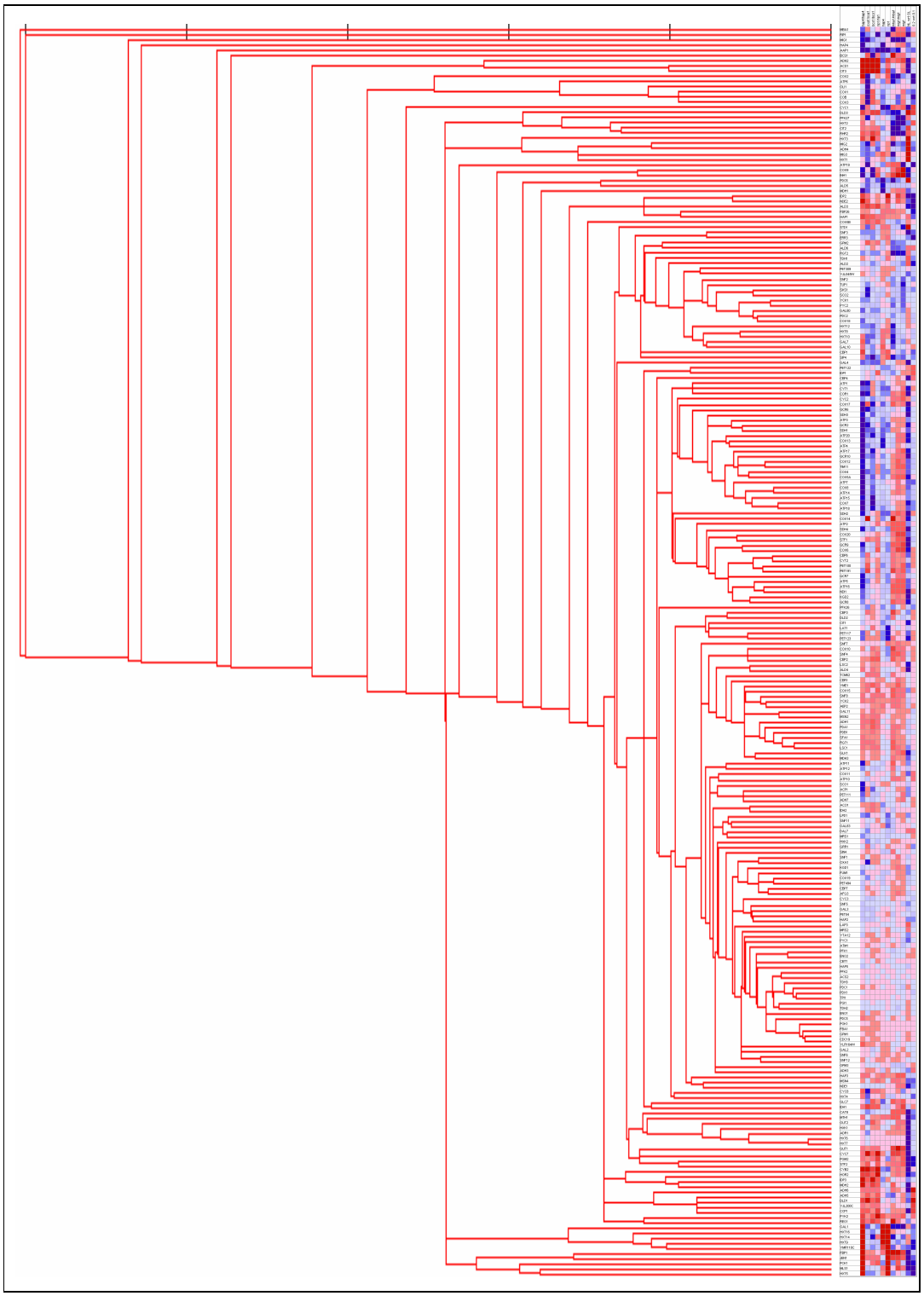

Figure 4.3. Hierarchical clustering of central carbon metabolism genes

Table 4.2. Distribution of the central metabolism genes on the clusters

Pathways (Genes)	Clusters (Genes)	0 (22)	1 (23)	2 (24)	3 (18)	4 (17)	5 (17)	6 (43)	7 (33)	8 (43)
Hexose transporters (13)		3	1	7	-	-	-	-	-	2
Glucose signaling pathway (18)		4	1	3	-	1	-	2	1	6
Glycolysis (24)		3	10	2	2	1	2	-	-	4
Galactose metabolism (11)		2	1	3	1	1	1	2	-	1
Glycerol metabolism (5)		-	2	-	-	-	1	-	-	2
Lactate metabolism (5)		-	1	-	-	2	-	-	-	2
Acetaldehyde from pyruvate (4)		2	2	-	-	-	-	-	-	-
Ethanol from acetaldehyde (8)		1	-	-	-	2	2	1	-	2
Acetate from acetaldehyde (5)		1	1	-	1	-	1	-	-	1
AcCoA from acetate (2)		-	-	-	-	-	1	-	-	1
AcCoA from pyruvate (5)		-	-	1	1	1	-	-	-	2
Glyoxylate cycle (5)		1	1	1	1	-	1	-	-	-
TCA cycle (18)		-	-	-	-	5	3	2	3	4
Cytochrome c related (5)		-	-	-	-	-	-	2	1	2
Complex I (4)		-	-	-	-	-	1	1	1	1
Complex II (8)		-	-	1	-	-	2	1	4	-
Complex III (10)		-	-	-	**1**	-	-	3	5	-
Regulation of Complex III (13)		1	-	1	-	1	-	1	5	2
Complex IV (12)		-	1	**1**	**2**	-	-	7	1	-
Regulation of Complex IV (19)		2	-	2	2	1	-	3	8	1
Complex V (18)		-	-	1	**3**	-	-	12	2	-
Regulation of Complex V (10)		-	-	-	1	1	-	2	2	4
Regulators of Central Metabolism (15)		2	1	1	2	-	2	1	-	6
Deleted genes (6)		-	1	-	1	1	-	3	-	-

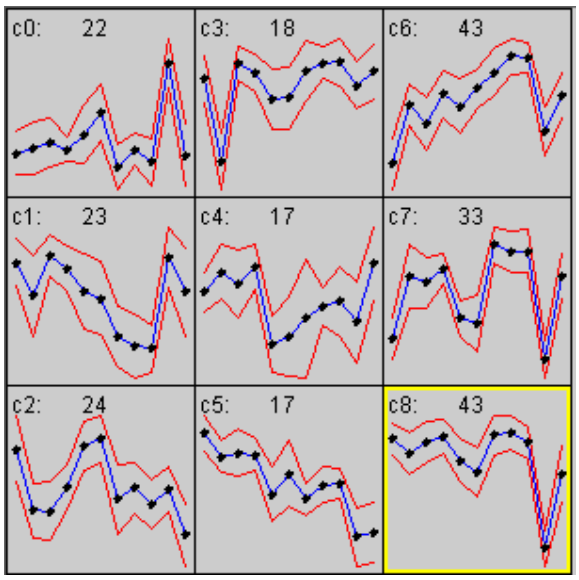

Figure 4.4. Centroids of clusters of central metabolism genes.

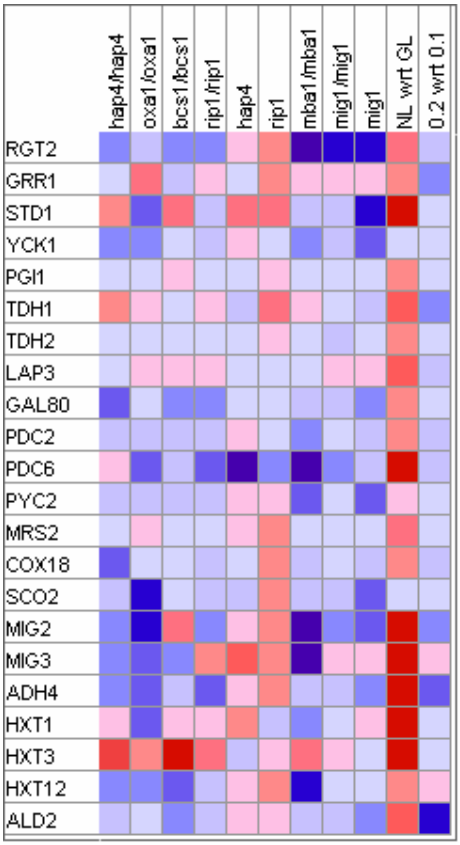

Figure 4.5. Analysis of expression levels of genes in Cluster 0

In Figure 4.5, the genes placed in Cluster 0 are shown, their common feature is being upregulated when glucose level is high. These genes encode three hexose transporters, four proteins from glucose signaling pathway, three enzymes of glycolysis, two pyruvate decarboxylases producing acetaldehyde, four regulators of galactose and central metabolism. It may be concluded that glucose repression caused by ammonium limited conditions activate these pathways, which generally lead to fermentation.

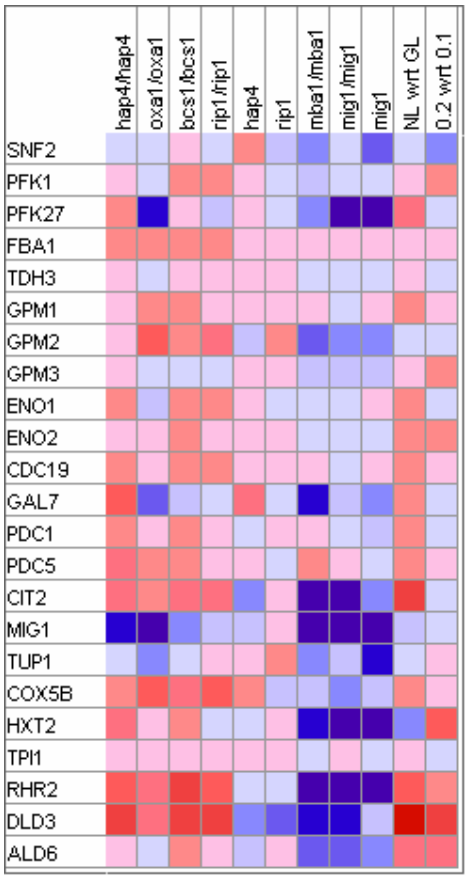

Figure 4.6. Analysis of expression levels of genes in Cluster 1

The genes shown in Figure 4.6 are placed in cluster and they are generally dowregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions. Also *RGT1* from Cluster 0 could have been placed in this cluster. Except for *MIG1*, the members of this cluster are generally upregulated in respiratory deficient strains (first four columns in Figure 4.6 represent the deletion that cause respiratory deficiency) and ammonium limitation. Genes encoding 10 out of 24 TCA cycle enzymes are placed in Cluster 1. Also genes encoding two pyruvate decarboxylases producing acetaldehyde and two enzymes in

glycerol metabolism are in this cluster. Generally this cluster indicates that the pathways leading to fermentation are upregulated in respiratory deficient mutants and in high glucose concentration (ammonium limitation) while they are dowregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1/mig1Δ* deletions.

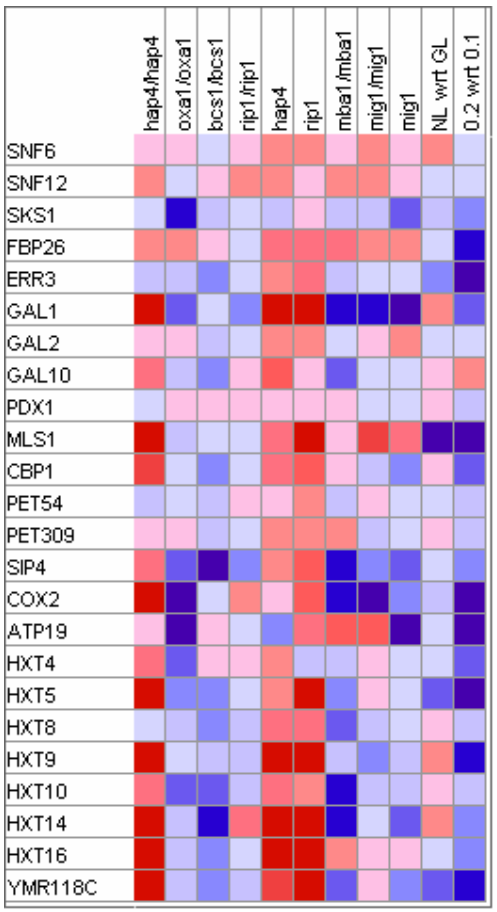

Figure 4.7. Analysis of expression levels of genes in Cluster 2.

In Figure 4.7, genes in Cluster 2 are shown. 7 of the 13 genes encoding hexose transporters are placed in this cluster. Also three genes from glucose signaling pathway, two genes from glycolysis, three genes from galactose metabolism and two genes encoding regulators of Complex IV are placed in this cluster. Expression profiles have an interesting motif in this cluster, *hap4Δ/hap4Δ, HAP4/hap4Δ* and *RIP1*/*rip1Δ* deletions upregulate the expression of these genes. Also most of them are downregulated at high growth rate. This behavior is referred to as "behavior of transporters" throughout this study, however no biological reasoning can be made.

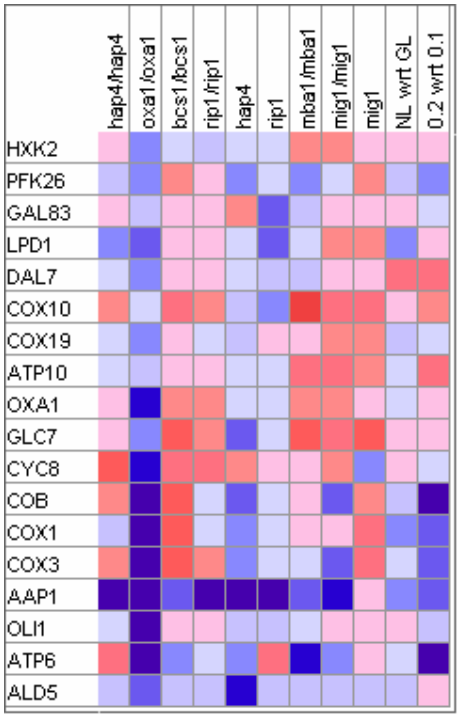

Figure 4.8. Analysis of expression levels of genes in Cluster 3

As it may be seen on Table 4.2 and Figure 4.8, six of the seven mitochondrially encoded genes in respiratory chain are placed in Cluster 3. These genes were strongly downregulated by *oxa1Δ/oxa1Δ* deletion. The other genes placed in this cluster are distributed to various pathways, five of them encodes either components or regulators of respiratory chains. Regulators *CYC8* and *GLC7* are also in this cluster together with *OXA1*.

Oxa1p is a protein involved in mitochondrial localization of proteins, however present results clearly indicate that it is involved in regulation of mitochondrially encoded respiratory chain components.

The genes placed in Cluster 4 are given in Figure 4.9. These genes are generally upregulated in respiratory deficient mutants and high growth rate. Products of five of these genes are involved in TCA cycle, catalyzing the reactions from citrate to α-ketoglutarate. These reactions in TCA cycle were previously reported to be induced in respiration deficiency to compensate the need for glutamate precursor α-ketoglutarate (Butow and Avadhani 2004). Also four enzymes from lactate metabolism and ethanol production pathways are placed in this cluster. Thus, it may be concluded that respiration deficiency

and high growth rate cause these induction of pathways leading to ethanol and to glutamate through α-ketoglutarate.

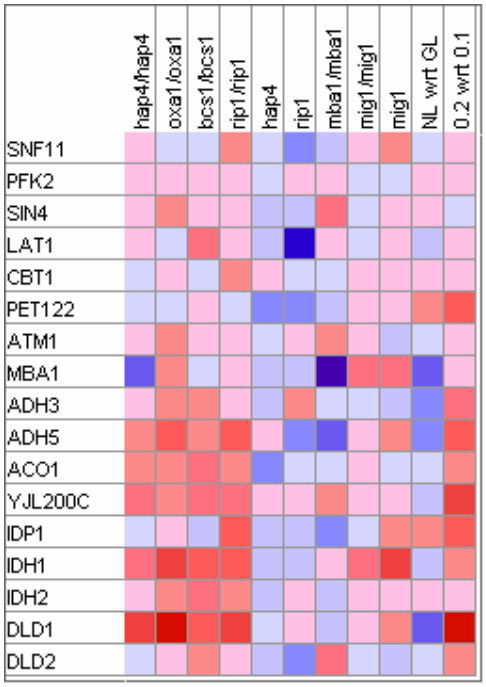

Figure 4.9. Analysis of expression levels of genes in Cluster 4

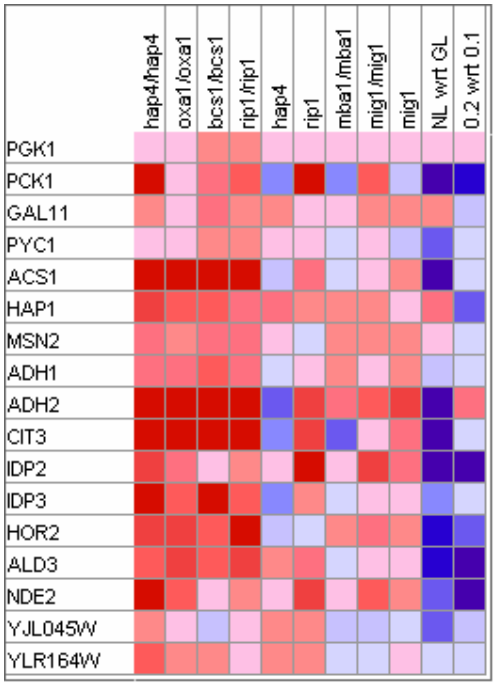

Figure 4.10. Analysis of expression levels of genes in Cluster 5

The genes from Cluster 5 (Figure 4.10) have a significant expression motif, these genes are upregulated by all gene deletions, downregulated when glucose level is high and high growth rate. These genes are distributed to many pathways like glycolysis, ethanol production, TCA cycle. This cluster is a variant of Cluster 4, differing only in responses to growth medium and growth rate.

In Figure 4.11, the genes placed in Cluster 6 are given. Genes in this crowded cluster are generally downregulated when glucose level is high, downregulated by *hap4Δ/hap4Δ* and *bcs1Δ/bcs1Δ* deletions while upregulated by *mba1Δ/mba1Δ*, *mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions. These genes encode either components or regulators of respiratory chain. The deleted genes *HAP4*, *BCS1* and *RIP1* are also placed in this cluster. Thus, it may be concluded that respiratory chain complexes are activated by *HAP4* and *BCS1* while being downregulated by high glucose levels, *MIG1* and *MBA1* genes. *BCS1* effect is not apparent for the genes shown in the upper part of Figure 4.11, these genes encode the regulators of the respiratory chain. Thus, probably *BCS1* is not effective on expression of the regulatory proteins but effective only on the components.

Cluster 7 given in Figure 4.12 is a variant of Cluster 6. Similarly, genes from respiratory chain dominate this cluster. Most of the genes given on the upper part (genes encoding the regulators) are not downregulated by *hap4Δ/hap4Δ* deletion and they are generally upregulated in respiratory deficient mutants. The genes located in bottom part of the cluster are components of the respiratory chain and downregulated by *hap4Δ/hap4Δ*, they are not significantly upregulated in respiratory deficient mutants. All genes in the cluster are upregulated by *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1*/*mig1Δ* deletions and downregulated when glucose level is high. Similar to Cluster 6, this cluster demonstrates that respiratory chain is activated by Hap4p, downregulated by glucose, Mig1p and Mba1p.

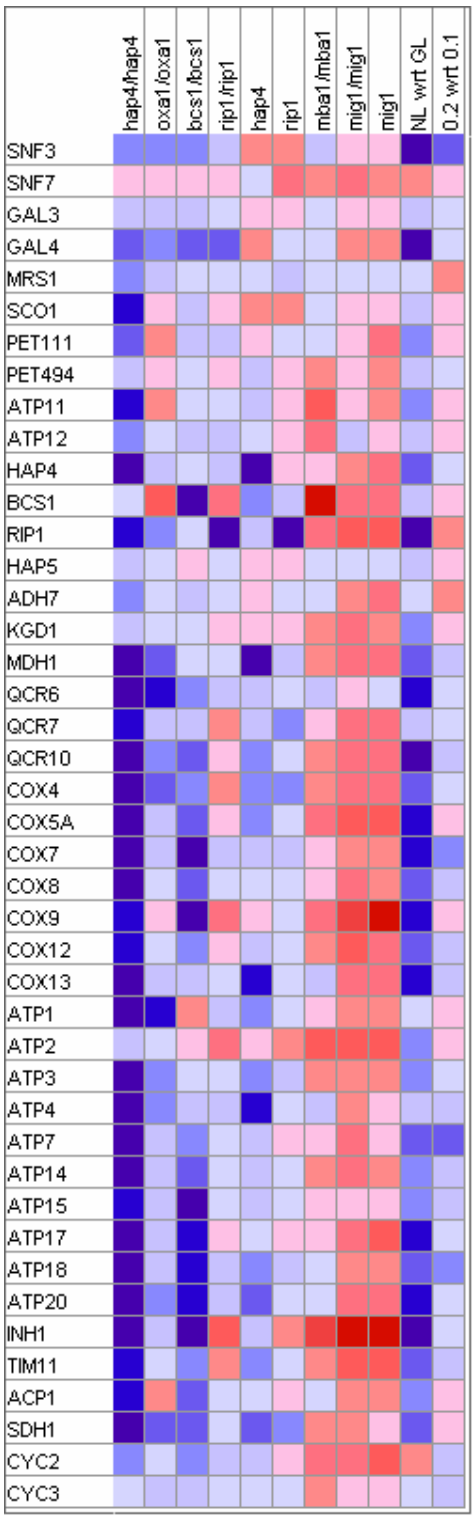

Figure 4.11. Analysis of expression levels of genes in Cluster 6

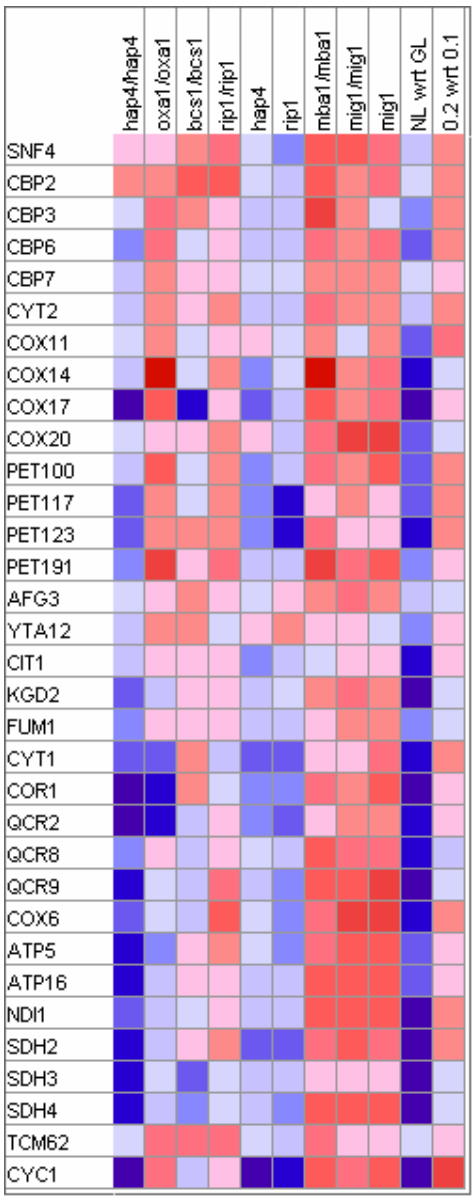

Figure 4.12. Analysis of expression levels of genes in Cluster 7

In Figure 4.13, genes in Cluster 8 are given. This crowded cluster contains the genes downregulated when glucose level is high and upregulated by almost all gene deletions. Regulatory proteins constitute a big group in this cluster. Almost all pathways have 1 to 6 members in this cluster except for respiratory chain complexes. Thus, all the central paths respond to respiratory deficiency, high glucose level, *mba1Δ/mba1Δ, mig1Δ/mig1Δ* and *MIG1/mig1Δ* deletions in the same direction: they are upregulated.

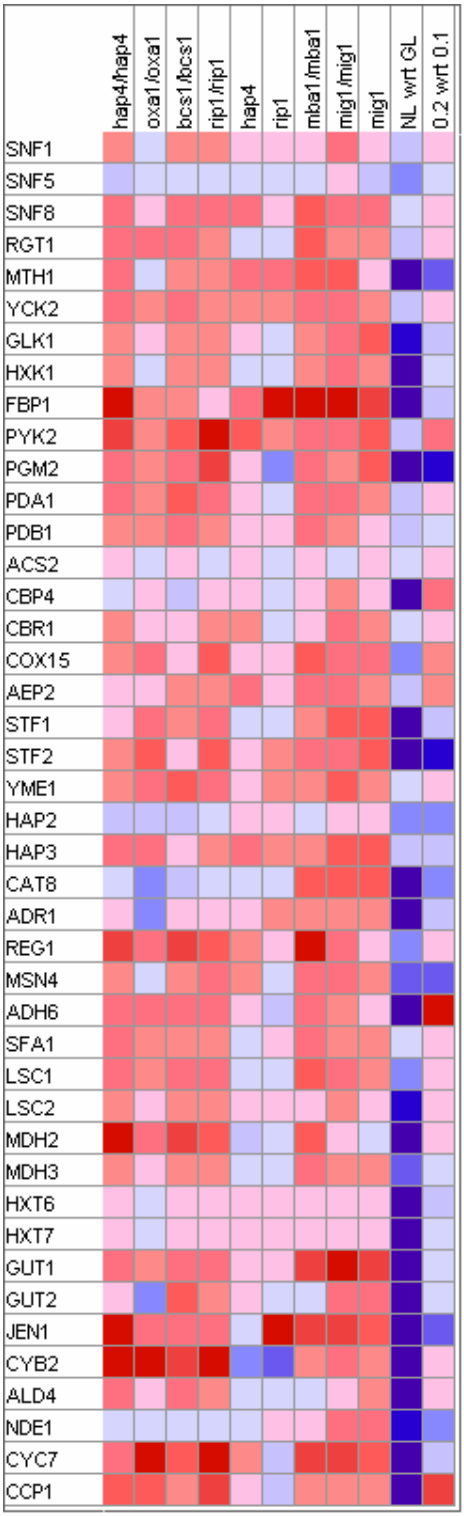

Figure 4.13. Analysis of expression levels of genes in Cluster 8

4.4. General Discussion of Effects of Deletions on Central Carbon Metabolism

The investigation of gene expression in deletion mutants resulted in interesting findings on regulation of central metabolism of yeast cells.

1. Most of the genes from respiration chain are downregulated by absence of *hap4Δ*/*hap4Δ* deletion. Thus, it is verified once more that Hap4p activates the genes from respiration chain. However, the genes involved in regulation of respiratory chain are not affected from *hap4Δ/hap4Δ* deletion.

2. Heterozygous deletion *HAP4*/*hap4Δ* generally causes a similar but weaker effect at the same direction when compared to the homozygous deletion *hap4Δ/hap4Δ*.

3. Generally, homozygous *rip1Δ/rip1Δ* deletion either has no effect or causes some induction in the expression of the other genes. The induction of a gene in *RIP1*/*rip1Δ* deletion mutant is generally accompanied by induction of that gene in the other respiratory deficient strains. Probably, Rip1p has no regulatory effect, however, it affects many genes through causing respiratory deficiency.

4. *bcs1Δ/bcs1Δ* deletion is effective on several components of Complex IV and V, but no pathway is generally effected from *bcs1Δ/bcs1Δ* deletion. *RIP1* expression is not affected from *bcs1Δ/bcs1Δ* deletion, thus Bcs1p regulation on *RIP1* should be post-transcriptional. Probably, Bcs1p has no regulation effect on transcription, but its deletion induces many genes since respiratory deficiency has global effects.

5. Complete respiratory deficiency upregulates many pathways from carbon central metabolism.

6. *mba1Δ/mba1Δ de*letion mutants behave similar to *mig1Δ/mig1Δ* deletion mutants. Heterozygous *MIG*1/*mig1Δ* deletion generally causes the same effect with the homozygous deletion.

7. Absence of *MIG1* and *MBA1* induces most of the genes in central pathways. Thus, Mig1p and Mba1p may be repressing both ethanol production and respiration. Only a few genes are strongly downregulated by their deletion.

8. Respiration deficiency induces many pathways, this may be an attempt of the cells to adapt to this severe deficiency. Glycolysis, glycerol metabolism, pathways from pyruvate to AcCoA, from acetaldehyde to ethanol, from acetaldehyde to acetate, acetate to AcCoA, TCA and glyoxylate cycles, regulation of Complex III and V are upregulated when the cells can not respire. Pathways producing ethanol and glycerol are strongly activated. First three steps of the TCA cycle are also activated to produce α-ketoglutarate, which is the precursor for glutamate production.

9. Ammonium limitation has a strong effect on central metabolism, probably because of the high glucose concentration in ammonium limited medium. The respiratory chain is generally downregulated while galactose metabolism and glycolysis are upregulated. Production of acetaldehyde from pyruvate is also upregulated which may indicate that glucose repression causes the carbon source to be utilized by fermentation.

10. *oxa1/oxa1Δ* deletion downregulates the mitochondrially encoded proteins of respiratory chain. Thus, Oxa1p may be a transcriptional activator of these genes.

11. The last point to be mentioned is the induction of some genes by *hap4Δ/hap4Δ, HAP4/hap4Δ* and *RIP1/rip1Δ* deletions. These genes are not affected from homozygous deletion *rip1Δ/rip1Δ*. They are mostly genes of hexose transporters, glucose signaling pathway, glycolysis and galactose pathways. Some genes from Complex I and II, and regulation of Complex III, also exhibit this interesting behavior.

5. INTEGRATION OF TRANSCRIPTOME AND METABOLOME

5.1. Modelling Expression Levels of ORFs and PCA

Linear modeling was used as a filtering tool to eliminate the ORFs with insignificant expression changes in response to the perturbation factors (growth medium, growth rate and gene deletion). Mean centered and scaled (unit variation) expression levels of 6361 ORFs were modelled and p-values were calculated to decide on the significance of the effects of the factors selected on the expression of the ORFs. For most of the ORFs, the constructed models did not predict a variation more significant than the expected level of random error, thus these ORFs were not included in further analyses. A relatively high p-value of 0.1 was used as the threshold, in order to include all ORFs that may possibly be affected by the three factors considered in the study.

638 out of 6361 models estimated that at least one of the factors was affecting the expression of the modelled ORF. The growth medium is the most effective factor in expression of most of the ORFs (56.3 per cent), followed by dilution rate (33.3 per cent), while gene deletion is the most effective factor for only 10.5 per cent of the ORFs.

The transcriptome data from all ORFs and the 638 ORFs were analysed separately using PCA. In Table 5.1, variations represented by the principle components (PC) are given in percentages. First three PCs represent most of the variation in filtered transcriptome data while more PCs should be included in the unfiltered data to represent more variance. The PCA of unfiltered data hardly distinguishes the samples from different conditions (Figures 5.1A and B). First PC distinguishes the samples from different media and second PC distinguishes the samples from different dilution rates (Figure 5.1C). Third PC distinguishes the samples from different deletion mutants (Figure 5.1D). These results prove that filtering of genes with insignificant response to perturbations via linear modelling removes most of the noise in transcriptome data.

Table 5.1. Per cent variation explained by PCs

Unfiltered Data			Filtered Data		
PC	Per cent variation	Cumulative per cent variation	PC	Per cent variation	Cumulative per cent variation
1	27.9	27.9	1	49.1	49.1
2	21.5	49.5	2	31.3	80.4
3	15.4	64.9	3	11.7	92.0
4	13.8	78.7	4	2.5	94.6
5	9.7	88.4	5	2.2	96.8
6	6.5	94.8	6	1.8	98.6
7	5.2	100.0	7	1.4	100.0

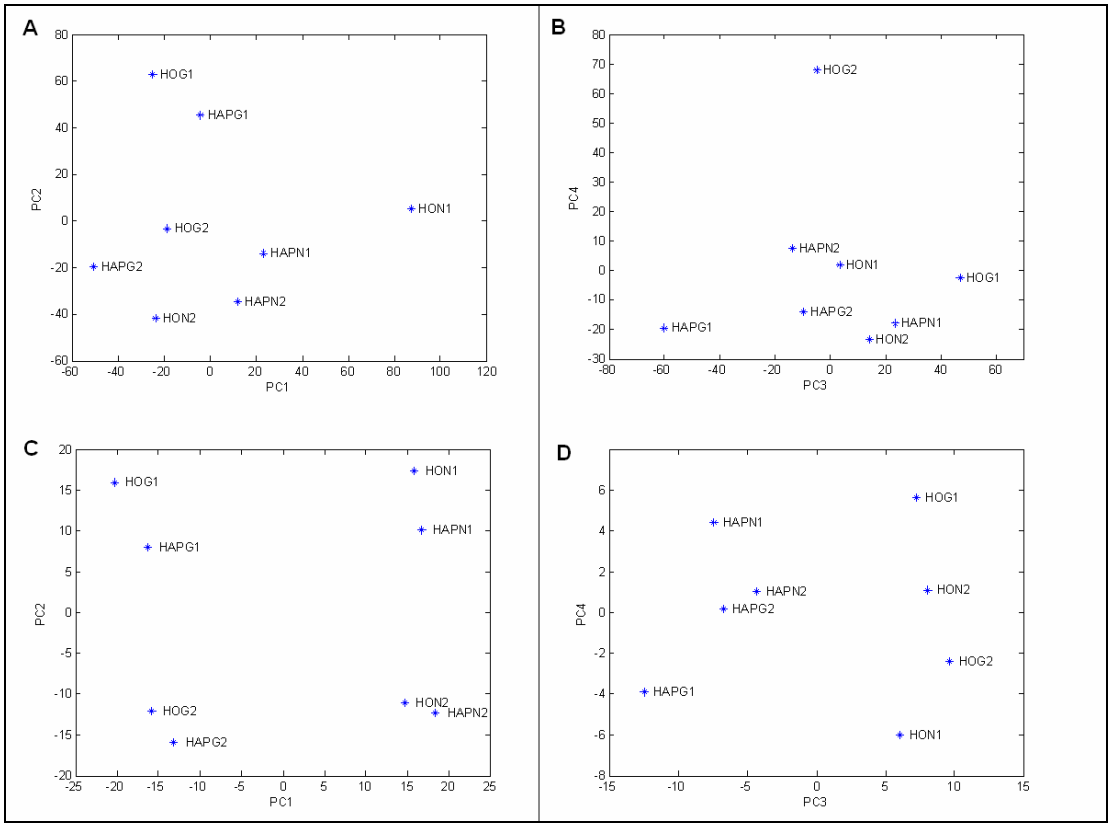

Figure 5.1. Scores of transcriptome data on first four PCs. A-B: PCA on complete transcriptome profile. C-D: PCA on filtered transcriptome profile.

5.2. Integration of Metabolic and Transcriptomic Data

The biomass concentrations obtained in chemostats operating at steady state are presented in Figure 5.2, together with the glucose consumption and ethanol production rates. The eight different conditions were selected on the basis of a factorial experimental design (Tables 3.9 and 3.10). The glucose consumption and ethanol production rates were found to increase with ammonium limitation, *hap4Δ*/*hap4Δ* deletion, and increasing dilution rate. Biomass concentrations, on the other hand, do not follow a similar trend; they decrease with increasing dilution rate. Under ammonium limitation, the biomass concentrations of *hap4Δ*/*hap4Δ* deletion mutants are only slightly higher than those of *hoΔ*/*hoΔ* deletion mutants.

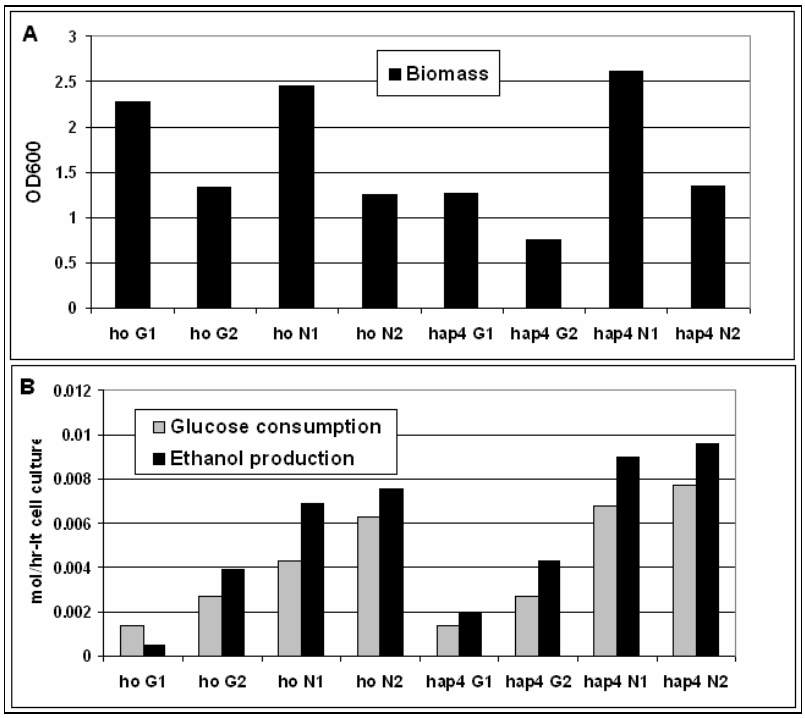

Figure 5.2. Biomass and metabolic data.

In order to model the metabolic data as a function of the transcriptome data, and to identify the ORFs that mediate the effects of the perturbations, partial least squares (PLS) method was applied to the transcriptome of 638 ORFs (X) and metabolic data (Y). Dimensions of the matrices X and Y were 8 x 638 and 8 x 3, respectively and both matrices were mean-centered and scaled to variance 1 prior to PLS regression. In Table

5.2, variations represented by the latent variables (LVs) are given in percentages. More than 80 per cent of the variations in both data sets are represented by the first two LVs.

Cumulative prediction error sum of squares (PRESS) are plotted to evaluate the prediction power and limitations of the constructed model (Figure5.3). A 'leave-one-out' procedure was used in PRESS calculations; *i.e.,* in each step, one of the samples was not included in the model and metabolic profile of the left-out sample was predicted using the model constructed; then the prediction error of that sample was calculated, and the procedure was repeated until all samples had been left out once.

The model with two LVs gave the minimum cumulative PRESS for all three response variables *i.e.*, biomass concentration, glucose consumption and ethanol production rates (Figure5.3). When the third LV was included, the cumulative PRESS increased, indicating that the prediction power of the model was not improved. However, the fourth LV did improve the model and explained some of the variance, therefore, LV3 and LV4 were included in the following interpretation. The observation of high biomass cumulative PRESS values for all the LVs indicated a low prediction power of the model for the biomass.

Table 5.2. Proportion of the variation explained by each latent variable

LV	Per cent variation (X)	Cumulative per cent variation (X)	Per cent variation (Y)	Cumulative per cent variation (Y)
1	48.9	48.9	63.3	63.3
2	31.3	80.2	28.5	91.8
3	7.9	88.2	3.7	95.5
4	6.1	94.2	3.2	98.7
5	2.0	96.3	1.2	99.9
6	2.3	98.5	0.1	100.0
7	1.5	100.0	0.0	100.0

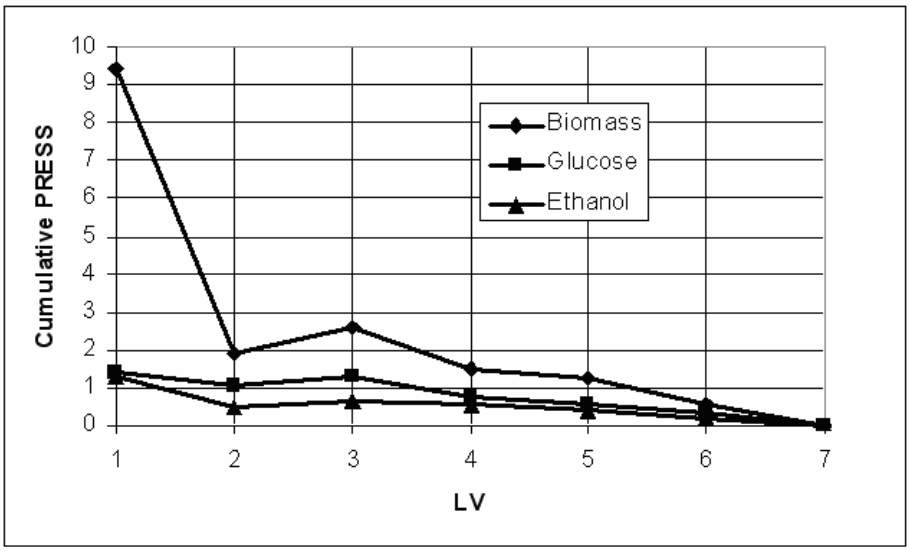

Figure 5.3. Cumulative prediction error sum of squares (PRESS) for biomass concentration, glucose consumption and ethanol production rates.

In order to see whether the distribution of the samples is linear according to Eq.3.23, the scores for the transcriptome (t) and the metabolome (u) on each LV were plotted against each other (Figure 5.4). Distribution of the scores around the fitted lines for LVs 1, 2 and 4 are quite good, thus the modeling capabilities of these LVs are satisfactory (Figures 5.4A, B and D, respectively), whereas the distribution of scores on LV3 is quite scattered, indicating a weakness of the prediction power of the model for variation of the metabolic data represented by LV3.

Scores of the transcriptome (t) and the metabolome (u) on the first four LVs are plotted in Figures 5.5 and 5.6, respectively. The first two LVs separate the samples taken from the different media and dilution rates (Figures 5.5A and Fig 5.6A). The highest variation (48.9 per cent of X and 63.3 per cent of Y), which is represented by LV1 in Table 5.2, is due to the medium factor and this is followed by the variation in LV2 (31.3 per cent of X and 28.5 per cent of Y) which is due to the dilution rate.

hap4Δ/hap4Δ deletion mutants resulted in positive scores on LV4 under all conditions (Figure5.5B and Figure5.6B). *hoΔ/hoΔ* deletion mutants, on the other hand, have negative scores on LV4, with the exception of one sample (*ho G1* in Figure 5.5B and

ho G2 in Figure 5.6B with slightly positive scores for the transcriptome and metabolome, respectively).

The samples from ammonium limitation score positively on LV1 (Figure 5.5A), and the response levels (Y values) under ammonium limitation are higher when compared to the glucose limited samples of the same mutant at an identical dilution rate, with only one exception.

Although the biomass has a positive loading on LV1, almost similar biomass concentrations obtained for *ho N2* and *ho G2* samples show that the biomass is not correctly predicted by the model for this particular case.

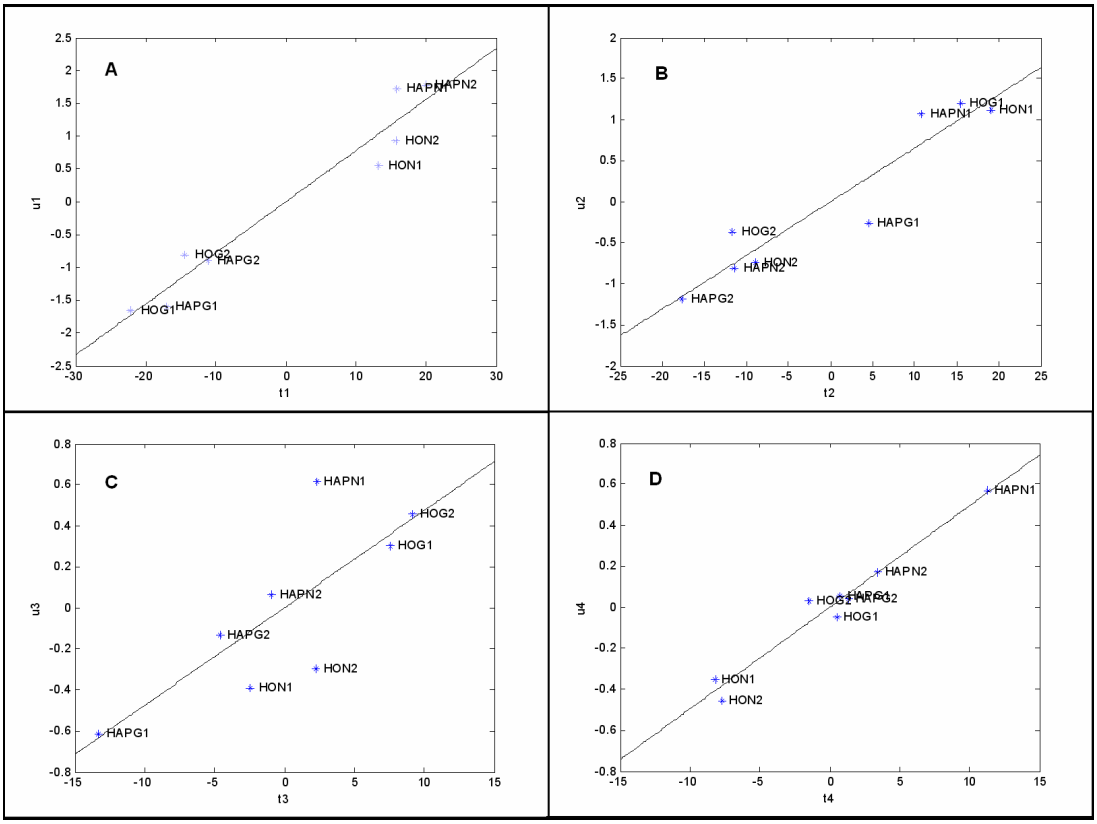

Figure 5.4. Comparison of scores for transcriptome and metabolic data.

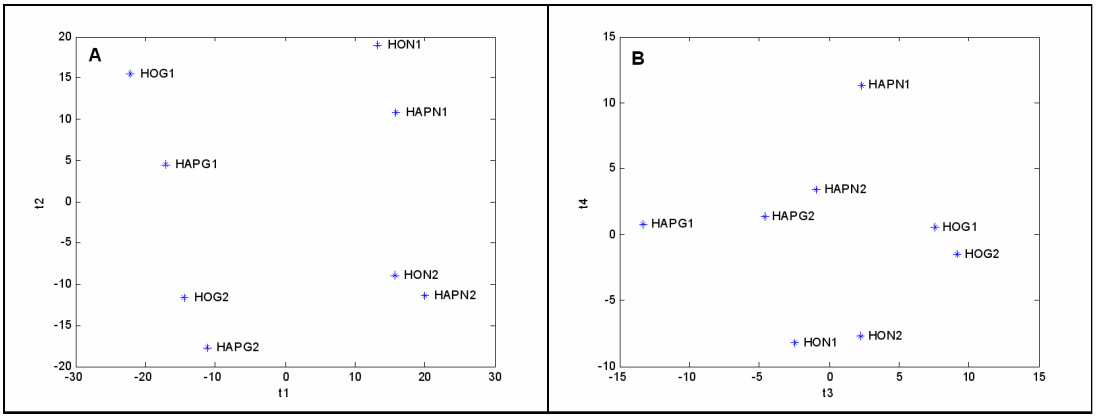

Figure 5.5. Scores of transcriptome data on first four LVs.

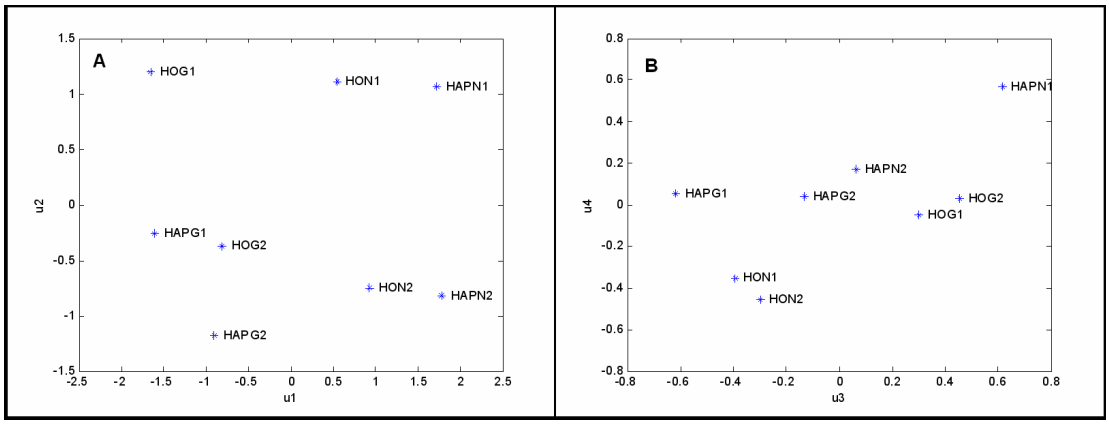

Figure 5.6. Scores of metabolic data on first four LVs.

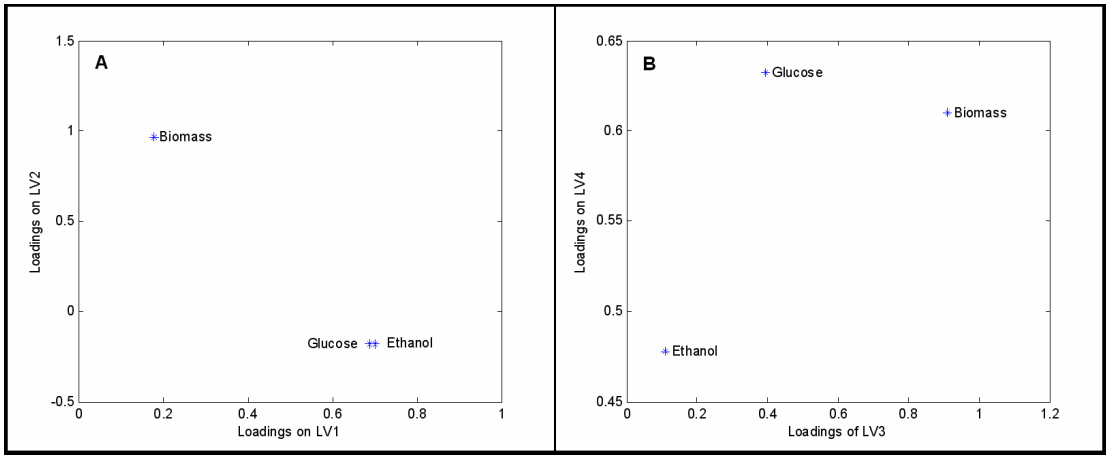

Figure 5.7. Loadings of metabolites on first four LVs.

In order to assess the contribution of biomass concentration, glucose consumption and ethanol production rates to the direction of the LVs, loadings of metabolites on the first four LVs are plotted in Figure 5.7. The behaviour of glucose consumption and ethanol production rates in response to medium and dilution rate factors were modeled by the first two LVs and their trends were found to be similar to each other (Figure 5.7A).

The ethanol and glucose load negatively on LV2, and they are affected negatively at the lower dilution rate (Figures 5.2 and 5.6A). The biomass has a positive score on LV2, and it is affected positively in the samples with positive scores on LV2, *i.e.* samples at the lower dilution rate. All response variables have positive loadings on LV4 and, therefore, would be expected to have higher values in the case of *hap4Δ*/*hap4Δ* deletion mutants, which have positive scores on LV4. However, this prediction does not hold for the biomass (Figures 5.2 and 5.6B). Thus, the model is successful in predicting metabolic data such as glucose consumption and ethanol production rates but fails for the biomass concentration beyond LV2.

5.3. Analysis of ORFs with significant contribution

Loadings of the ORFs on the latent variables were investigated to unravel the relationship between the transcriptome and response variables (Figures 5.8 and 5.9). The ORFs with positive loadings on an LV are up-regulated in samples with positive scores on that LV while they are down-regulated in samples with negative scores (see Table 3.9 for sample names). On the other hand, the ORFs with negative loadings on an LV are down-regulated in samples with positive scores on that LV.

Variance in LV1 represents the differences due to the medium factor. The genes with positive loadings on LV1, which are up-regulated under ammonium limitation compared to carbon limitation, are expected to be the genes that mediate the increase in biomass concentration, ethanol production rate and glucose consumption rate. Similarly, the genes with negative loadings on LV1 are most likely to be the genes that are up-regulated at glucose limitation causing the decrease in these response variables.

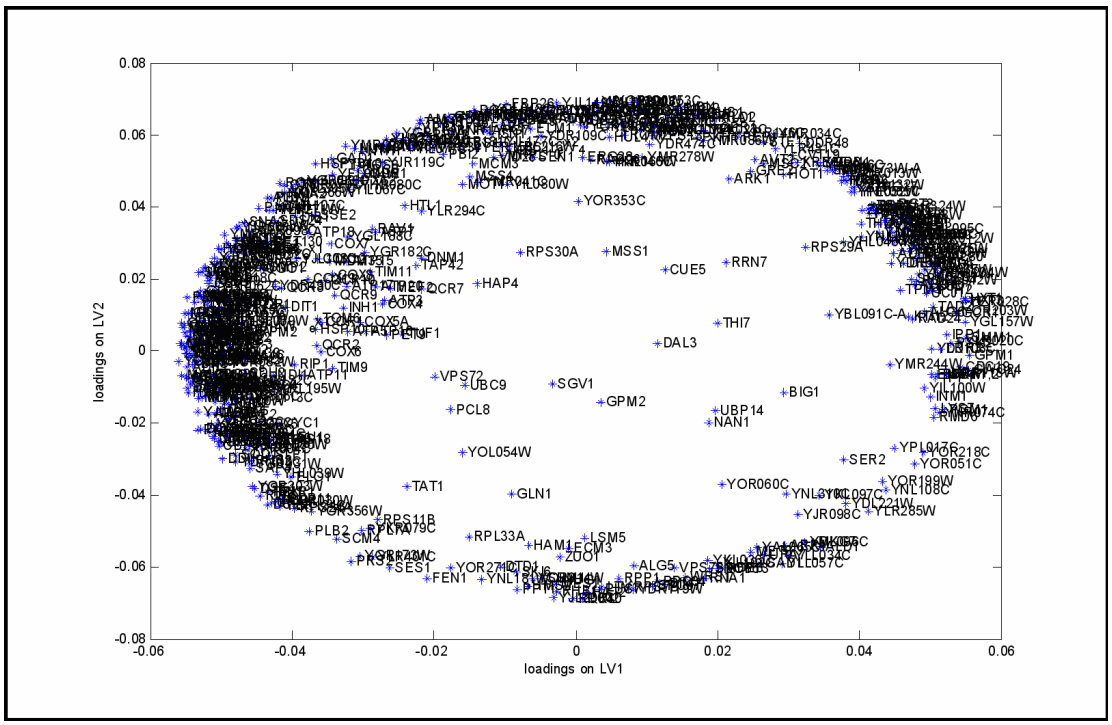

Figure 5.8. Loadings of ORFs on first and second LVs.

The ORFs with positive loadings on LV2 are up-regulated at lower dilution rate since samples from the lower dilution rate have positive scores on LV2, and the increase in biomass concentration as well as the decrease in ethanol production and glucose consumption rates are expected to be mediated by these ORFs. Similar analysis on LV4 indicates that ORFs with negative loadings on LV4 are down-regulated in *hap4Δ*/*hap4Δ* deletion mutants, possibly affecting glucose consumption and ethanol production rates.

Student's t-test was applied to loadings of the ORFs, and only those ORFs with $t < 1 \times 10^{-6}$ are listed in the Table 5.3. GO Mapping was applied to these ORFs and only the significant biological process terms at the 6th level of the hierarchy were included to represent the groups of similar terms. (+) and (–) signs indicate positive and negative loadings of the ORFs, respectively.

The ORFs that are down-regulated under ammonium limitation as compared to glucose limitation are the genes that are active in the polysaccharide metabolism, ribonucleotide metabolism and respiration pathways (LV1-). The glucose concentration in

the ammonium-limited medium should repress the expression of genes acting on the metabolism of other sugars and in the respiratory pathways. Repression of respiration, in turn, would cause the fermentation pathway to be activated and ethanol production to be enhanced. Hexose transport is activated in glucose limited case, probably for effective transport of the scarce source of carbon.

The ORFs up-regulated at the lower dilution rate (LV2+), are the genes that act on oxidative phophorylation. Up-regulation of these ORFs mediates the increase in biomass and the decreases in ethanol production and glucose consumption rates.

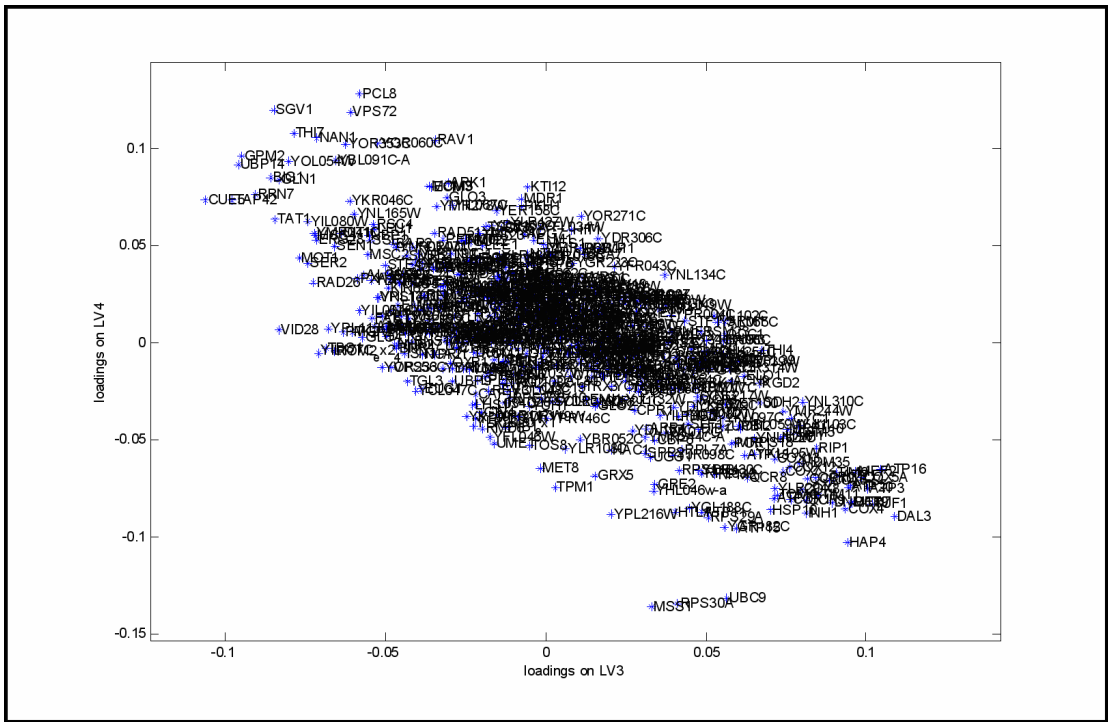

Figure 5.9. Loadings of ORFs on third and fourth LVs.

The genes that were down-regulated due to *hap4Δ*/*hap4Δ* deletion (LV4-) are mainly respiration-related, *i.e.*, genes that function in oxidative phosphorylation, cofactor metabolism and ion transport. Consequently, the *hap4Δ*/*hap4Δ* deletion causes respiratory deficiency under the conditions studied, and fermentation is the only process of glucose metabolism. The higher glucose consumption and ethanol production rates achieved

provide further confirmation that Hap4p plays a major role in the switch mechanism from respiration to fermentation.

Biomass production is a complicated process which cannot be simply attributed to a single metabolic pathway and, in this work, modelling of the biomass concentration was successful only for the dilution rate factor. On the other hand, modelling of the ethanol production and glucose consumption rates were successful for all the factors considered; and these response variables were found to be a function of the expression of the genes that are actually acting on the pathways related to their metabolism.

The number of genes given in groups LV1-, LV3+, and LV4- (Table 5.3) that are members of the GO terms "cellular respiration", "phosphorylation", "coenzyme biosynthesis" and "nucleotide metabolism" have high significance ($p < 1.0E^{-10}$). The unknown ORFs that appear in the same group with the above mentioned genes may also be members of the functional categories denoted by the over-represented GO terms.

Table 5.3. ORFs and GO terms with highest contributions to the LVs

LV	ORFs with significant loadings	Cellular Process	P-value
LV1+	ALD1, BUB3, HXT1, HXT3, MNT4, MSC2, YER028C, YGL157W, YJL132W, ZRT2	hexose transport	0.0003
LV1-	AAC1, ACN9, ADR1, ATP16, ATP17, ATP18,ATP3, ATP5, BAP2, BNS1, CBP4, COX12, COX4, COX6, COX7, COX8, EMI2, GAC1, GDB1, GLG1, GPH1, GSY1, HOR2, HXK1, HYP2, INH1, ISF1, KGD2, MAL31, MAL33, MBR1, MCR1, MRK1, MTH1, PCL7, PIG1, PRS2, PRX1, QCR10, QCR8,	phosphorylation cellular respiration ribonucleotide metabolism cellular polysaccharide metabolism coenzyme biosynthesis	2.63E-14 1.85E-10 1.72E-09 3.36E-07 4.27E-07

Table 5.3. continued

	QCR9, ROM1, STF2, UIP4, YDL157C, YFL043C,YFR017C, YGR243W, YJL103C, YJL185C, YJR008W, YKL187C, YLR327C, YMR090W, YMR103C, YNL122C, YOL053W, YPL230W, YPR196W		
LV2+	ALD2, AMS1, ATP17, BUB3, CHS1, CNM67, COX12, COX7, COX8, CPR1, FBP26, GAD1, GPG1, HOR2, HSP78, HXT1, MRPL10, MSC2, PHM8, PIG1, PRX1, RAD2, ROM1, UBI4, YBL112C, YFL043C, YJL103C, YJL132W, YJR008W, YMR034C, YMR090W, YOL153C, YOR220W, YPL216W, YPR117W, ZRT2	oxidative phosphorylation	8.01E-05
LV2-	ALD1, ARR4, BAP2, CBP4, MRK1, PCL7, PRS2, RPC40, RPL7A, YLL034C, YOR271C, YOR314W	regulation of catabolism	0.00031
LV3+	ACN9, ATP15, ATP16, ATP17, ATP18, ATP20, ATP3, ATP5, ATP7, CBP4, COR1, COX12, COX4, COX5A, COX6, COX7, COX8, CYC1, *HAP4*, HSP78, HXT1, INH1, KGD2, MCR1, MRPL10, MRPS18, PRS2, PRX1, QCR10, QCR8, QCR9, RPL7A, TUF1, YDL157C, YGL157W, YJL103C, YNL122C, YOR314W	phosphorylation coenzyme biosynthesis nucleotide metabolism cellular respiration ion transport	4.43E-21 1.45E-11 2.83E-11 3.16E-11 4.45E-09
LV3-	BAP2, CNM67, ISF1, MRK1, MSC2, PIG1, ROM1, TAP42, TAT1, UBP14, YFL043C, YGR243W, ZRT2	regulation of carbohydrate biosynthesis	0.00037

Table 5.3. continued

		amino acid transport energy derivation by oxidation of organic compounds carbohydrate biosynthesis	0.00194 0.00437 0.00499
LV4+	BAP2, GDB1, GLG1, GLO3, HOR2, MAL33, MRK1, MSC2, PIG1, UBP14, YBL091C, YOR271C, YPR196W	carbohydrate metabolism energy derivation by oxidation of organic compounds glucan metabolism	1.06E-06 1.41E-05 4.23E-05
LV4-	ATP15, ATP16, ATP17, ATP18, ATP20, ATP3, ATP5, ATP7, CBP4, COR1, COX12, COX4, COX5A, COX6, COX7, COX8, *HAP4*, HXT1, INH1, MCR1, MRPL10, QCR10, QCR8, QCR9, RPL7A, TUF1, YDL157C, YJL103C, YJL132W, YNL122C, YPL216W	nucleotide metabolism cellular respiration cofactor metabolism ion transport	1.05E-10 1.15E-10 3.81E-10 6.06E-10

5.4. Analysis of Metabolome and Transcriptome of Deletion Mutants

In this section effects of gene deletions on metabolome and transcriptome are discussed and modelling of metabolome profile using transcriptome data are presented.

5.4.1. Metabolite Levels in Deletion Mutants

The effects of the gene deletions on biomass production, glucose consumption and ethanol production rates were investigated in chemostats The results are presented in Figure 5.10. 20 - 40 per cent reductions in biomass production were observed in deletion mutants that have respiratory deficiency (*hap4Δ/hap4Δ*, *bcs1Δ/bcs1Δ*, *oxa1Δ/oxa1Δ* and

rip1Δ/rip1Δ mutants). The strains that are partially respiratory deficient (*HAP4*/*hap4Δ*, *RIP1*/*rip1Δ*, and *mba1Δ/mba1Δ*) did not have any growth deficiency. Glucose consumption of the cells remained the same, indicating that respiratory deficient cells consumed more glucose per biomass when compared to others. The ethanol production did not follow a particular trend among the mutants.

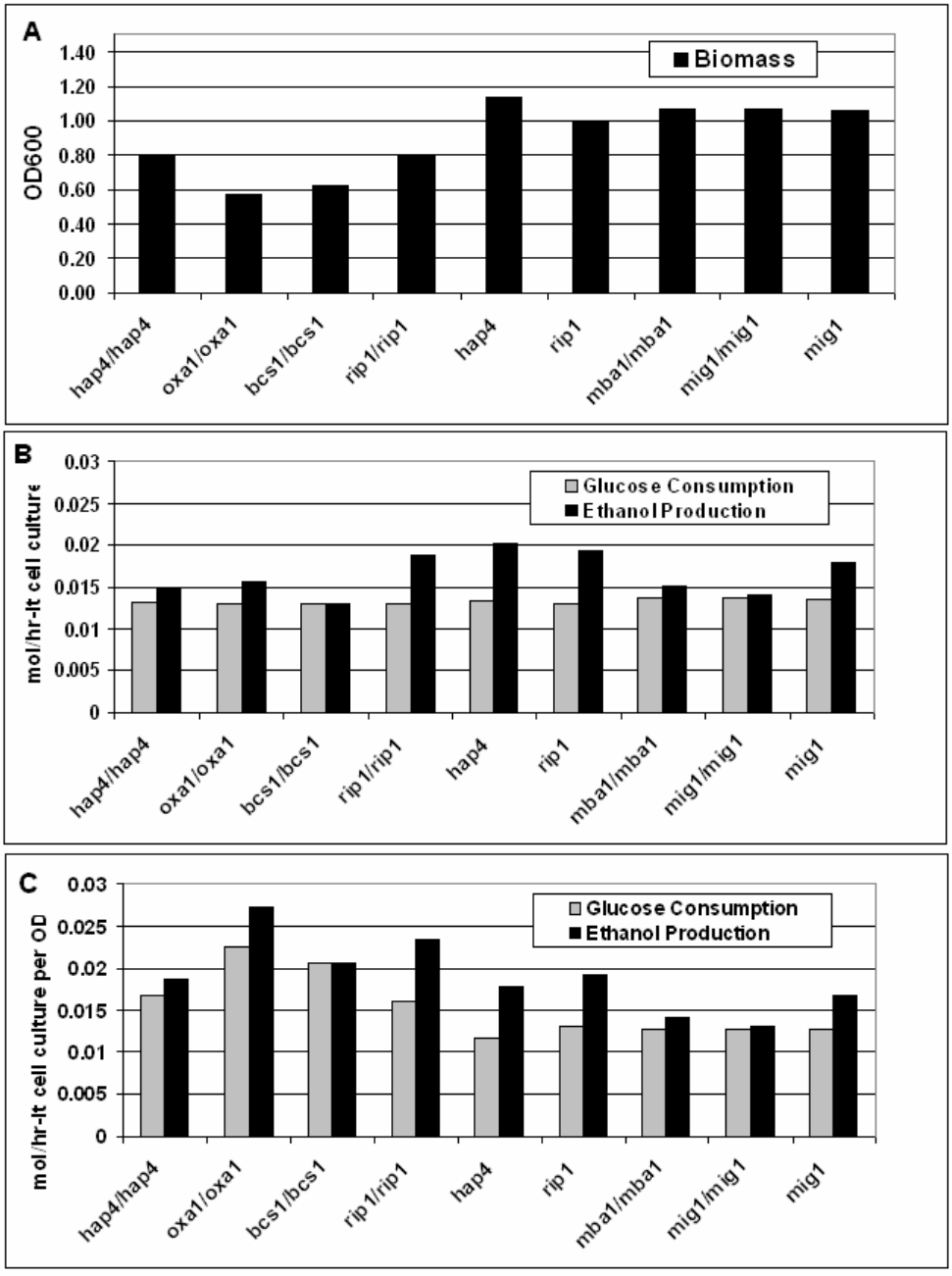

Figure 5.10. Biomass and metabolic data in various deletion mutants.

The ethanol glucose concentrations per OD are plotted in Figure 5.10C to compare their consumption and production rates per biomass production respectively. Glucose consumption rate is significantly higher in respiratory deficient cells, indicating that respiratory deficiency causes ineffective utilization of glucose in terms of biomass production. Ethanol production rate per biomass is also higher in these cells, which partly explains the inefficient utilization of glucose in respiratory deficient cells for biomass and energy. The pyruvate produced in glycolysis should be used in fermentative metabolism since respiration does not occur in these cells. Thus, ethanol is produced as an end product of glucose metabolism and growth defect is observed since precursors of biomass production can not be produced.

5.4.2. GO Terms Affected from Deletions

The analysis of the genes significantly affected from deletions were performed. The effects of deletions on the gene expression were evaluated using linear modelling. A total of 2163 ORFs were found to be affected by at least one of the deletions significantly ($p < 0.1$), only homozygous gene deletions were considered. The biological process terms overrepresented among the upregulated and downregulated genes are plotted in Figure 5.11. The biological processes overrepresented by genes downregulated by a deletion are assumed to be activated by the deleted gene and vice versa. HAP4 activates, *i.e.*, *hap4Δ/hap4Δ* downregulates, many processes generally related to respiration, homeostasis, nucleotide and ribosome metabolisms. Some of these processes were also downregulated by other gene deletions causing respiratory deficiency and upregulated by *mig1Δ/mig1Δ* mutant. The processes upregulated by *hap4Δ/hap4Δ* (given as "downregulated by HAP4" in Figure 5.11) are generally carbohydrate, nitrogen, organic acid, alcohol metabolisms and response to stress. These processes are also downregulated by other genes that cause respiratory deficiency when deleted from the genome.

On the network, the genes whose deletion affect common GO terms appear close to each other. In respiratory deficient mutants of *hap4Δ/hap4Δ*, *bcs1Δ/bcs1Δ*, *oxa1Δ/oxa1Δ* and *rip1Δ/rip1Δ*, similar biological processes are affected, while some of those processes are inversely affected from *mig1Δ/mig1Δ* and *mba1Δ/mba1Δ* deletions.

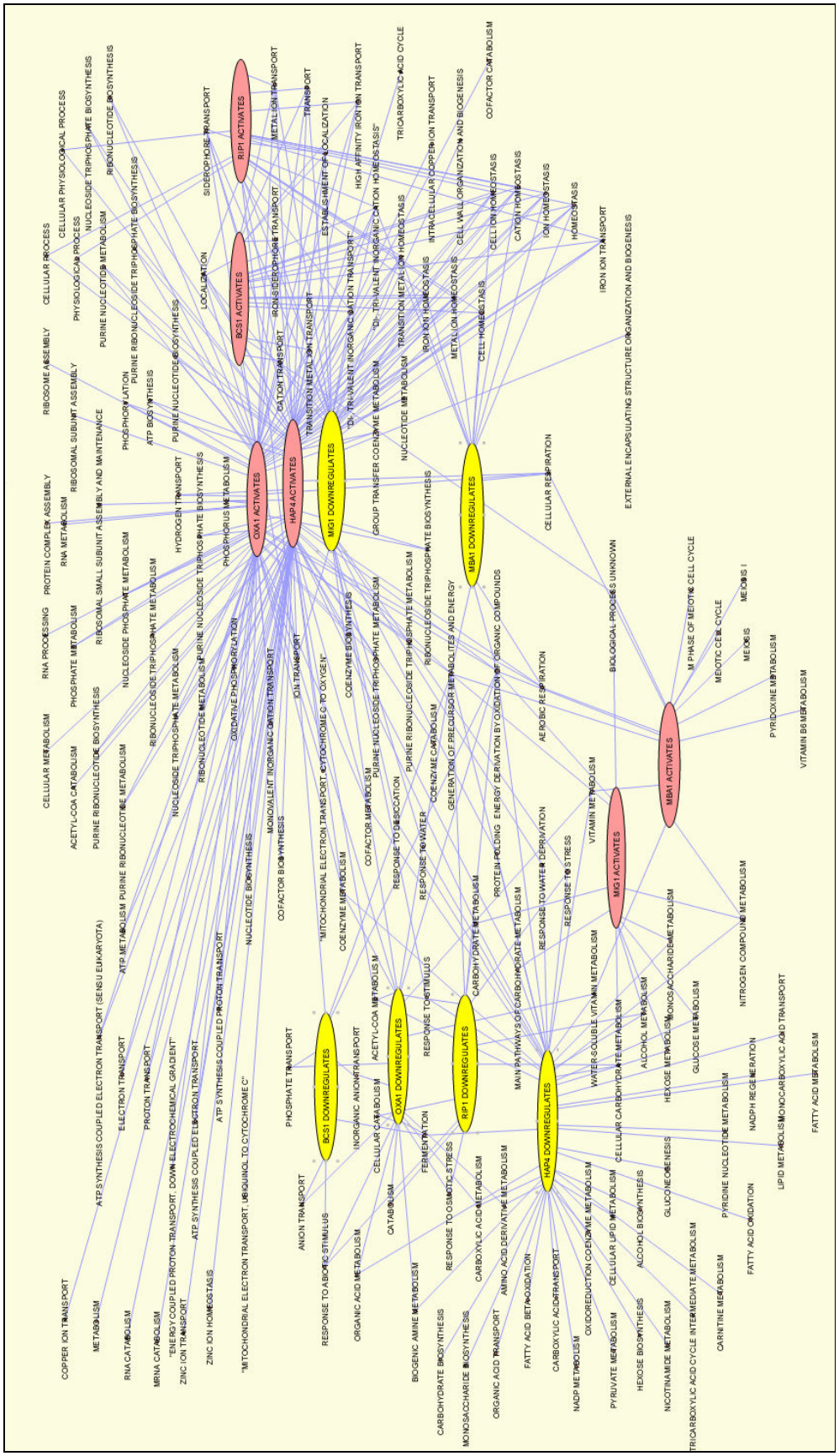

Figure 5.11. GO terms overrepresented by significantly affected genes

5.4.3. PCA of Transcriptome Data from Deletion Mutants

PCA was applied to autoscaled expression data of 2163 significantly affected ORFs, percent variations explained by the principle components are given in Table 5.4. The first two PCs distinguish between the replicates. The deletion mutants *mig1Δ/mig1Δ* *mba1Δ/mba1Δ* and *MIG1/mig1Δ* are clustered together while other homozygous deletion mutants are scattered as pairs on the projections given in Figure 5.12. Heterozygous deletion mutants *HAP4*/*hap4Δ* and *RIP1*/*rip1Δ* are distant from the other mutants. Third and fourth PCs distinguish *oxa1Δ/oxa1Δ* mutant from the others but other mutants do not appear as pairs of replicates.

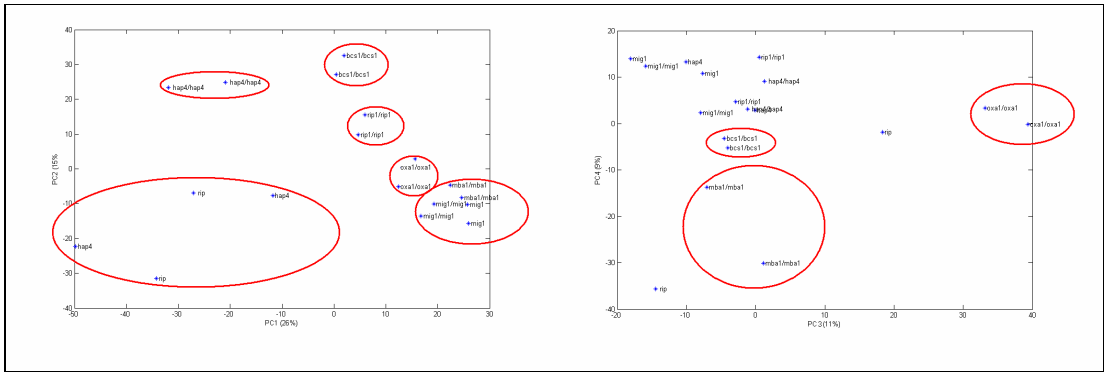

Figure 5.12. Scores of filtered transcriptome data on first four PCs.

Table 5.4. Percent variation explained by PCs

PC	Per cent Variation	Cumulative per cent variation
1	27.9	27.9
2	21.5	49.5
3	15.4	64.9
4	13.8	78.7
5	9.7	88.4
6	6.5	94.8
7	5.2	100.0

5.4.4. PLS of Transcriptome and Metabolites Related to Deletion Mutants

PLS was applied to metabolites (biomass, glucose consumption and ethanol production rates) and to filtered transcriptome data (2163 genes) after averaging the replicate measurements. First latent variable discriminates the samples from respiratory deficient and respiratory sufficient strains (Figure 5.13). The strains with partial respiratory deficiency cluster with the strains which have no respiratory deficiency. The latent variables other than the first one do not contribute to the model, i.e. less than 0.2 per cent of variance is represented by the other LVs (Table 5.5).

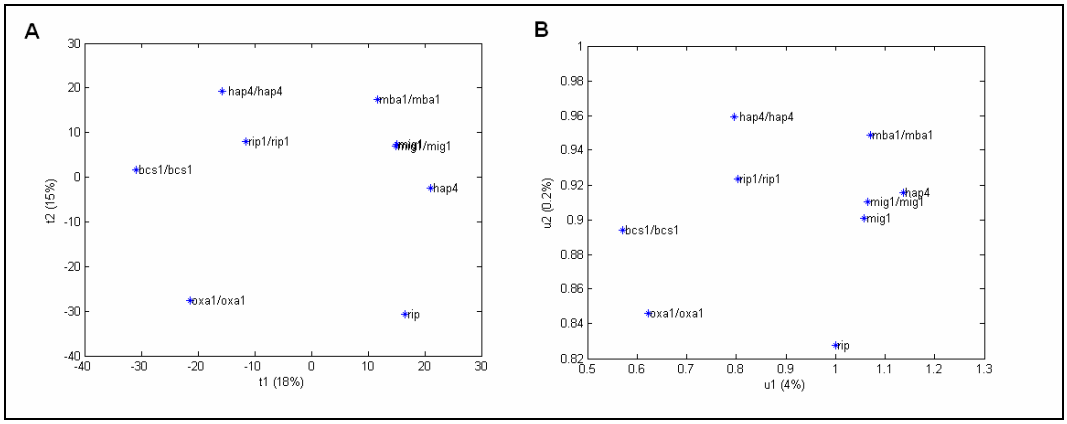

Figure 5.13. Scores of transcriptome (A) and metabolic data (B) on LV1 and LV2.

Table 5.5. Proportion of the variation explained by each latent variable

LV	Per cent variation (X)	Cumulative per cent variation (X)	Per cent variation (Y)	Cumulative per cent variation (Y)
1	17.9	17.9	4.4	4.4
2	14.7	32.6	0.2	4.6
3	26.5	59.0	0.0	4.6
4	14.0	73.0	0.0	4.6
5	7.7	80.7	0.0	4.6
6	6.4	87.1	0.0	4.6
7	7.1	94.2	0.0	4.6
8	5.8	100.0	0.0	4.6
9	0.0	100.0	95.4	100.0

Table 5.6. ORFs and GO terms with highest contributions to the LVs

LV	ORFs with significant loadings	Cellular Process	P-value
LV1+	YDR018C, YHL035C, YBR250W, YCL002C, YIL166C, YGL072C, YPR195C, YJL132W, YMR251W, YGR102C, YCR099C, YMR206W, YIR043C, YLR137W, YOL155C, YJR079W, YMR110C, YER158C, APT2, ARN1, ARN2, ATX1, CCC2, COT1, CTL1, CTR2, DCD1, DUO1, ENB1, FET3, FIT2, FIT3, FRE1, FRE3, FRE5, FTR1, GAL4, GPI2, MAL11, MEK1, MTM1, RCE1, RRP8, SIT1, SMF3, STV1, TIM18, TIS11, YGP1, ZRT3	ion transport metal ion homeostasis	2.4E-14 4.5E-10
LV1-	YIR007W, YGL157W, ACS1, ADH2, CWP1, CYB2, DIP5, GDH2, GIP2, IDP3, MCT1, MSC1, POX1, SCM4, SED1, SPS19, SSU1, TSA2, RNR2, TFS1, PHO8, LYS1, PEX21, PHO84, YIL083C, PDR17, GCY1, TIP1, GAD1, PHO89, GRX3, HOR2, ALD3, CTT1, HXT3, UBP15, YJL217W, SNO1, ASN1, SSE2, YHR202W, CRC1, TRP3, YNL179C, GLN1, YOR285W, RTS3, QDR3, EMP47, ISA1	organic acid metabolism fatty acid metabolism fermentation	8.9E-8 1.2E-5 1.9E-4

The top 50 ORFs with positive and negative loadings on LV1 are given in Table 5.6. The biological processes overrepresented among these ORFs are investigated using GO Mapping, and the terms at 6^{th} level of hierarchy with low p-values are given.

ORFs involved in organic acid metabolism and fermentation are the ORFs that are active in respiratory deficient strains; while activation in ion transport pathways

discriminates the respiratory sufficient strains from the deficient ones. These findings are in accordance with the analysis given in Figure 5.11, which illustrated the biological processes affected from each gene deletion.

5.4.5. Integration of Transcriptome and Metabolome Data by PLS

Finally, the ORFs which are effective on complete the metabolome profile of respiratory deficient and sufficient cells are analysed. The intracellular and extracellular metabolite samples are analysed with MS, and the binned MS spectra are modelled using the transcriptome data via PLS.

The discrimination of the samples is not possible when the intracellular metabolome profile is analysed via PLS (results not shown); however, extracellular metabolite profile can be modelled successfully. The percent variations explained by LVs are given in Table 5.7, and the scores of the samples on first two LVs are given in Figure 5.14.

Table 5.7. Proportion of the variation explained by each latent variable

LV	Per cent variation (X)	Cumulative per cent variation (X)	Per cent variation (Y)	Cumulative per cent variation (Y)
1	25.0	25.0	9.4	9.4
2	18.4	43.3	3.3	12.6
3	17.0	60.4	1.8	14.5
4	9.6	69.9	1.1	15.5
5	12.1	82.1	0.5	16.0
6	7.0	89.0	0.6	16.6
7	4.1	93.2	0.3	16.9
8	6.8	100.0	0.1	17.0
9	0.0	100.0	83.0	100.0

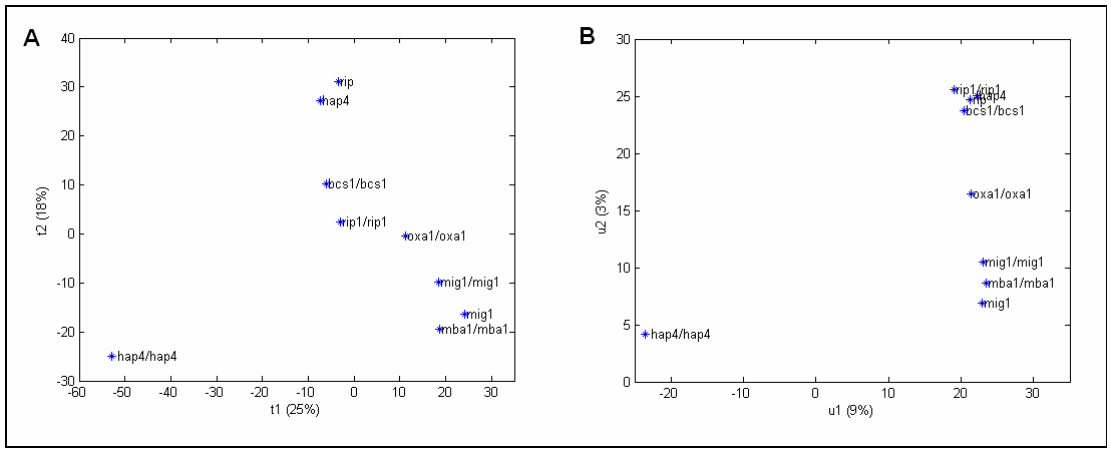

Figure 5.14. Scores of transcriptome (A) and extracellular metabolome (B) on LV1 and LV2.

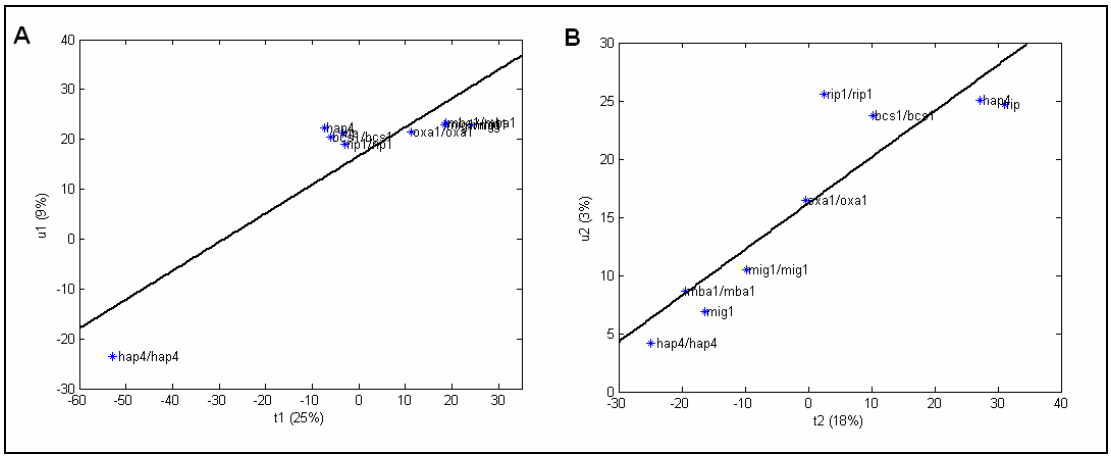

Figure 5.15. Comparison of scores for transcriptome and extracellular metabolome data.

The scores of transcriptome and extracellular metabolome given in Figure 5.14 show that homozygous deletion *hap4Δ/hap4Δ* mutant may be an outlier among the other deletion strains. The heterozygous deletion mutants *RIP1*/*rip1Δ* and *HAP4*/*hap4Δ* cluster with the other respiratory deficient strains. Figure 5.15 shows the comparison of scores for transcriptome and extracellular metabolome data on LV1 (Figure 5.15A) and LV2 (Figure 5.15B). The modelling capability of LV2 is better when compared to LV1. LV1 explains the variation due to difference between *hap4Δ/hap4Δ* strain and other deletion mutants; then, LV2 explains the differences among the other strains. The loadings of the ORFs and MS spectra are given in Figure 5.16. In Figure 5.16A, each point corresponds to loadings

of an ORF significantly affected from at least one deletion. In Figure 5.16B, each point corresponds to loadings of a peak observed between 20-1700 m/z. The peaks appearing in the spectra can not be assigned to a particular metabolite, thus the ORFs that would model each metabolite can not be detected. However, it is possible to comment on the ORFs which generally model the extracellular metabolites. In Table 5.8, the top 50 ORFs mediating the effect of gene deletions on metabolome profiles are given with the GO biological process terms overrepresented among these ORFs. The GO analysis indicates that the ORFs acting on oxidative phosphorylation are the main source of variation between *hap4Δ/hap4Δ* strain and other strains. Similarly, the ORFs acting on regulation of metabolism are found to be the source of variance among all strains.

Table 5.8. ORFs and GO terms with highest contributions to the LVs

LV	ORFs with significant loadings	Cellular Process	P-value
LV1+	HAP4, TIM11, SNO2, YMR166C, MRPL1, LEU2, COX7, QCR10, SDH1, YPL183W-A, CYC1, COX12, MHT1, MRPL3, RRN5, YKR077W, YKL066W, ATP14, MTO1, YDR493W, YMR003W, ATP18, MRPL13, COX8, IMG1, PTP1, YLR168C, COX17, HEM2, YPR011C, ATP7, YCL057c-a, ATP11, REC107, YLR364W, RSM25, MRP49, PDS1, YDR084C, SRX1, COX5A, YJL051W, YBL059W, POR2, ATP3, FMC1, MRPL25, CPR3, TUF1, SDH3	oxidative phosphorylation aerobic respiration nucleotide biosynthesis	1.9E-14 1.1E-7 3.9E-5
LV1-	YHL005C, ERG6, YOL047C, YPL272C, YHR126C, ERG1, YPL035C, YDR360W, TKL2, ATG23, CTA1, YGL036W, YIL055C, YDL094C, PHM8, TKL2, YOR387C, MUM2, YGR131W, YEH1, YBL054W, HXT2, YDR070C, SDS24, PRM5, ECM23,	steroid metabolism	2.6E-4

Table 5.8. continued

	YLR012C, ROG1, YGL258W, JLP2, YBR219C, RTA1, PEX27, YLR164W, YBR178W, YDL218W, HES1, AIR1, YOR287C, YDR140W, PXA1, YKR106W, YHL041W, PPR1, HBT1, YEL010W, YHL039W, VID24, YNL194C, FAA2		
LV2+	YPR090W, YLR057W, YJR020W, YLR143W, YGR198W, BUD14, BUD8, BUD9, CAF4, EAP1, EBS1, EDE1, GAL80, GLN3, GZF3, HAP5, HIR3, HMS2, IFH1, INP53, KRE11, MCM3, MCX1, MDS3, MIP1, MKC7, MLP1, MSG5, NAT1, NDC1, NMA2, NUP116, PHO91, PMD1, PTK2, PYC2, RAD24, RAV1, RCK2, RUD3, SCT1, SEC16, SEN1, SGV1, SOG2, TSC11, VPS13, YCK1, YCK3, ZMS1	regulation of metabolism	2.6E-5
LV2-	YKL187C, YML131W, ARC18, RUB1, SCS2, GPX2, ZEO1, GPI11, FLO9, PPM2, SSP1, CWP2, YBR047W, GPI15, GIM4, YDL193W, YOR215C, MAD2, YOR052C, YLR414C, RBK1, NMD4, PET10, RNR4, YBR281C, YOR220W, DER1, TSA1, YNR064C, YLR281C, YDL121C, DOT5, AAP1', SLM4, YIM1, RPN9, MPM1, HHO1, ERP4, PRX1, SMB1, APA2, YER030W, DTD1, YBR071W, MXR1, YBR014C, YIR016W, RPC40, YMR009W	cell redox homeostasis response to oxidative stress	4.8E-5 6.1E-3

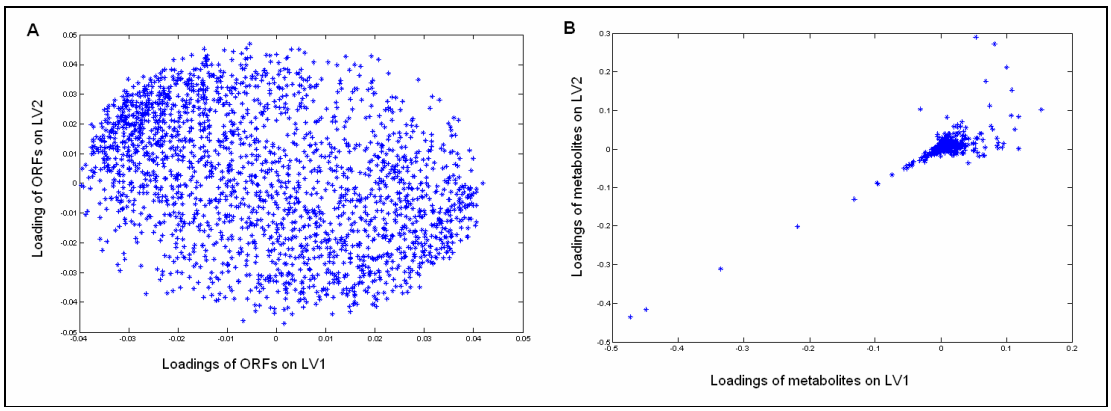

Figure 5.16. Loadings of ORFs and extracellular metabolites on LV1 and LV2.

6. ANNOTATION OF UNKNOWN YEAST ORFS

Linear modeling of mean-centered and scaled (unit variation) expression levels of 6361 ORFs was carried out as a filtering tool to eliminate ORFs that showed insignificant expression changes in response to the perturbations. 638 out of 6361 models predicted that at least one of the factors was affecting the expression of the modeled ORF ($p < 0.1$). The ORFs and p-values for the factors are given in Table B.1. ORFs whose transcription changed significantly in at least two of the experimental data sets (Table 6.1 and Table B.2) were selected. These ORFS were analysed using GO Slim Mapping to elucidate the biological processes that are affected from the perturbations applied in these experiments.

Table 6.1. Experiments and conditions

Exp.	Source	Conditions	Reported Genes
1	Present work	Glucose vs. Ammonium limitation	P< 0.01 in models of gene expression
1	Present work	*hap4Δ/hap4Δ* deletion vs. *hoΔ/hoΔ* deletion	P< 0.1 in models of gene expression
1	Present work	Dilution rate 0.2 vs 0.1hr^{-2}	P< 0.1 in models of gene expression
2	Hayes *et al.* (2002)	Carbon vs. Nitrogen limitation (D = 0.2 hr^{-1})	Top 17 Genes induced
3	Boer *et al.* (2003)	Carbon vs. Nitrogen limitation (D = 0.1 hr^{-1})	2-fold change from others
4	Buschlen *et al.* (2003)	*hap4Δ* deletion vs. WT (batch, complete galactose medium)	Down-regulated by deletion
5	Wu *et al.* (2004)	Glucose vs. Ammonium starvation (D = 0.1 hr^{-1})	Top 50 genes induced
6	Tai *et al.* (2005)	Carbon vs. Nitrogen limitation (D = 0.1 hr^{-1})	2-fold change from others
7	Grotkjaer (2005)	Dilution rates from 0.02 to 0.33	Clusters of up- and down-regulated ORFs

The ORFs whose expression profiles correlated with those of affected ORFs with unknown function were selected and analysed with GO Term Finder. The significantly over-represented biological processes among the correlated genes were annotated to the unknown genes under investigation. A literature survey was carried out to discover any further evidence supporting the proposed annotations. The annotations proposed were also compared with the previously published results of automated annotation methods. The present methodology was validated by re-annotating the known genes and evaluating the fraction of correct re-annotations made.

6.1. Effect of Factors on Biological Processes

Among the affected 638 genes, those up-regulated or down-regulated in nitrogen-limited conditions when compared to glucose-limited conditions, or in dilution rate $0.2hr^{-1}$ when compared to $0.1hr^{-1}$, in at least two of the experiments were selected (Table B.3). GO Slim (Dolinski *et al*., 2005) analysis was applied to classify the 227 ORFs according to their biological processes. Similarly, genes down-regulated in a *hap4Δ/hap4Δ* mutant, both in present study and in that of Buschlen *et al*. (2003), were selected and classified by GO Slim Mapping.

GO Slim terms that yielded $p < 0.01$ for at least one condition are given in bold in Table 6.2. The abbreviation UN-DC denotes upregulation at nitrogen limitation wrt carbon limitation, DN-UC denotes downregulation at nitrogen limitation wrt carbon limitation, DH denotes downregulation at *hap4Δ/hap4Δ* deletion, UR2-DR1 denotes upregulation at higher dilution rates wrt lower dilution rates, and DR2-UR1 denotes downregulation at higher dilution rates wrt lower dilution rates. Numbers in paranthesis give the number of genes. Details of these findings are given in Tables B.2, B.4 and B.5. The total number of genes (out of 6274) annotated to each term is given in parentheses. The total number of genes that were significantly affected at the investigated conditions and numbers of the affected genes annotated to each GO Slim term are given with the corresponding p-values. Of the genes up-regulated in carbon-limited conditions, 9/45 were annotated to carbohydrate metabolism which involves 219 genes, p-value is $6.73x10^{-6}$. Similarly, 6/47 genes up-regulated at the lower dilution rate were annotated to carbohydrate metabolism with a p-value of $2.77x10^{-3}$. Thus, significantly large number of genes involved in

carbohydrate metabolism were up-regulated under carbon limitation and at lower dilution rates. Catabolite repression is eliminated under carbon limitation which, in turn, up-regulates the carbohydrate metabolism pathways. At the lower dilution rate, the growth rate is lower and a greater propportion of available resources are used in carbohydrate metabolism to generate energy and precursor metabolites.

Table 6.2. Classification of ORFs by GO Slim Mapping ($p < 0.01$)

GO Slim Term and Number of Genes Annotated to Terms	P_{UN-DC}	P_{DN-UC}	P_{DH}	$P_{UR2-DR1}$	$P_{DR2-UR1}$
carbohydrate metabolism (219/7274)	> 0.01 (3/47)	**6.73E-06** (9/45)	> 0.01 (0/29)	> 0.01 (0/36)	**2.77E-03** (6/47)
cellular respiration (88/7274)	> 0.01 (0/47)	> 0.01 (2/45)	**6.74E-18** (13/29)	> 0.01 (0/36)	> 0.01 (2/47)
electron transport (33/7274)	> 0.01 (0/47)	> 0.01 (0/45)	**6.84E-17** (10/29)	> 0.01 (0/36)	> 0.01 (0/47)
generation of precursor metabolites and energy (226/7274)	> 0.01 (1/47)	> 0.01 (5/45)	**8.35E-30** (23/29)	> 0.01 (0/36)	**9.23E-05** (8/47)
organelle organization and biogenesis (1009/7274)	> 0.01 (2/47)	> 0.01 (1/45)	> 0.01 (1/29)	**7.73E-03** (11/36)	> 0.01 (2/47)
ribosome biogenesis and assembly (254/7274)	> 0.01 (0/47)	> 0.01 (0/45)	> 0.01 (0/29)	**3.08E-06** (9/36)	> 0.01 (0/47)
RNA metabolism (537/7274)	> 0.01 (0/47)	> 0.01 (1/45)	> 0.01 (0/29)	**4.10E-03** (8/36)	> 0.01 (0/47)
transport (899/7274)	**3.71E-03** (13/47)	> 0.01 (11/45)	**1.77E-03** (10/29)	> 0.01 (6/36)	> 0.01 (5/47)
biological process unknown (2468/7274)	> 0.01 (11/47)	> 0.01 (12/45)	> 0.01 (1/29)	> 0.01 (7/36)	> 0.01 (20/47)

However, at high growth rate, cells allocate more resources to organelle organization/biogenesis, ribosome biogenesis/ assembly and RNA metabolism in order to adapt to the higher growth rate (11, 9 and 8 genes out of 36 upregulated genes correspondingly). The *hap4Δ/hap4Δ* deletion causes respiratory deficiency, thus the pathways acting on respiration, electron transport and energy generation are deactivated by down-regulating the genes (13, 10, 23 out of 29 down-regulated genes, correspondingly) involved in these pathways. Genes acting on transport processes were down-regulated under carbon-limited and *hap4Δ/hap4Δ* deletion conditions (13 out of 47 and 10 out of 29, correspondingly) showing that these perturbations cause an adaptation in transport pathways at the transcriptional level. The GO Slim terms that were represented with $p > 0.01$ were excluded from the list; the terms may share some genes.

Expression of the genes *DUR1,2, DAL1, DAL2, DCG1, DAL7, ECM38, DAL3* which are all up-regulated under nitrogen limitation were mapped as "others" in GO Slim Mapping (Supp. Table 4). For further analysis of the genes in this group, GO Term Finder was used. 5/7 genes were found to be members of allantoin (an intermediate or end product of purine catabolism) metabolism which has total of 7 annotations. Four genes (*DUR1,2, DAL1, DAL2, DAL3*) have hydrolase activity, acting on nitrogen-carbon bonds (but not on peptide bonds). 6/7 genes are members of nitrogen compound metabolism, with the exception of *ECM38*. Though not apparent in GO Slim Mapping, nitrogen compound metabolism is significantly up-regulated under nitrogen-limited conditions.

6.5. Unknown ORFs Responding to Perturbations

Application of this analysis has also revealed 50 unknown ORFs that are affected by the perturbations in at least two of the experiments included in this study. Functional annotation can be made for these ORFs by GO Slim mapping. However, such an analysis does not provide adequate specificity. In order to annotate a specific term to these unknown ORFs, correlation coefficients between their expression profiles and the 6361 ORFs in the data set were calculated. The ORFs that are correlated ($\rho > 0.95$) with each of the unknown ORFs were grouped and GO mapping of the members of each group was performed (Table 6.3). Significant GO terms ($p < 0.001$) were found for more than one

third of the unknown ORFs (18/50). Results of the GO Term mapping for the genes in each group are given in Table B.6 for each significantly affected unknown ORF.

Table 6.3. Gene Groups Correlated to Unknown ORFs

Unknown ORFs	Genes Correlated ($\rho > 0.95$)	
	Number	Names
YMR090w	12	*GDH3, UGA2, YBR116C, HXT5, YHR199C, YKL107W, YGR043C, RAD7, HSP104, GAD1, YMR323W, YPR151C*
YGL157w	9	*YAL037W, CDC19, PTH2, YER028C, DUR3, HXT1, HXT3, YLR020C, SMM1*
YGR243w	30	*MAL33, MAL31, GLK1, YER067W, GOS1, GUT1, YIL057C, MGA2, MBR1, JEN1, GSY1, MRK1, AHA1, TRS23, MTH1, EMI2, PYC1, POP6, YGR067C, BNS1, PIG1, YLR327C, AAC1, ISF1, CAT8, YNL144C, GAC1, YPL230W, GAL4, GDB1*
YLR327c	27	*MAL32, GLC3, YER067W, CPR2, XBP1, MBR1, JEN1, GLG1, GSY1, YFR017C, YDR018C, YAP6, MTH1, STF2, YGR067C, BNS1, YGR243W, MAL11, YLR149C, PIG1, AAC1, ISF1, PGM2, YNL200C, GAC1, YPL230W, GPH1*
YER121w	22	*YBR062C, MAL32, YIL055C, XBP1, MBR1, GLG1, YFR017C, TSC13, YDR018C, YAP6, MTH1, AGA2, SIP2, STF2, YGR067C, PIG1, YLR327C, AAC1, ISF1, YNR014W, GAC1, GDB1*
YFR017c	31	*PRX1, YBR062C, ARA1, GLC3, HOR2, YER121W, CPR2, YIL055C, GTT1, MBR1, YDR018C, YAP6, MTH1, SIP2, STF2, YGR067C, YJL185C, YJR008W, PIG1, YLR327C, YLR408C, ISF1, YMR206W, YNL200C,GCY1,YOR173W,YOR285W,UIP4,OXR1,YPR098C, SNA2*
YGR067c	33	*YBR053C, YBR062C, MAL32, HOR2, YER121W, GOS1, CPR2, YIL055C, GTT1, MBR1, JEN1, GLG1, YFR017C, TSC13, YDR018C, YAP6, MTH1, EMI2, STF2, YGR243W, YJL045W, YJL185C, PIG1, YLR327C, YLR408C, ISF1, YMR206W, YNL200C, GCY1, GAC1, UIP4, CSR2, GDB1*
YKL187c	16	*BAP2, MAL33, MAL31, HYP2, GIP2, DOG2, MGA2, YFL052W, GSY1, HXK1, MRK1, PYC1, RNR2, MMP1, SPS19, ALD4*
YGR236c (SPG1)	9	*SSA3,ADY2,SPS100,LRS4,SIP18,YMR191W,MLS1,YNL195C, GRE1*
YMR107w(SPG4)	12	*PTC4, SLI15, GOS1, GTT1, PPZ2, PRP3, EMI2, YGL185C, RPS14B, VTI1, YNL200C, YOL032W*
YMR206w	19	*SLI15, YBR194W, YER067W, CPR2, MBR1, YFR017C, AHA1, YAP6, MTH1, STF2, YGR067C, PIG1, YLR408C, CAT8, GCY1, YOR179C, LGE1, CSR2, ISA2*
YER067w	21	*MAL31, MAL32, MBR1, CMK1, MRK1, AHA1, MTH1, BNS1, YGR243W, PIG1, YLR327C, AAC1, ISF1, YMR206W, CAT8, YNL144C, GLO4, YPL230W, GAL4, CSR2, GDB1*
YJL103c	16	*ACH1, GRX1, MDH1, COX20, ATP17, COX13, QCR9, QCR10, QCR8, YJR080C, COX12, COX8, COX7, POR1, MDM35, ATP18*
YNL175C(NOP13)	2	*LOC1, IMP4*
YJL200C	7	*RHB1, HAT2, GCS1, SXM1, PPT1, RPA190, RPC40*
YDR070C (FMP16)	16	*YBR116C, PBN1, ADY2, YKL107W, STP4, NDE2, AFR1, PMC1, ALD3, SIP18, YMR191W, YNL195C, YNL305C, YOR097C, YOR366W, GRE1*
YGR173W(RBG2)	6	*PRS2, DED81, SES1, SNF4, YLR401C, APT1*
YOR173W (DCS2)	15	*YBR149W, SDS24, HOR2, UBP9, YFR017C, YGR066C, ROM1, YJR008W, YLR271W,TSL1, YMR103C, PGM2, PHM7, ATH1, SNA2*

A detailed survey was conducted to provide additional evidence for specifying the function of unknown ORFs. For this purpose, the available literature as well as SGD and the Comprehensive Yeast Genome Database (CYGD, Güldener *et al.,* 2005) were screened for: motifs present in amino-acid sequences of their products, expression profiles in various growth conditions, phenotypes of deletion mutants, physical and genetic interactions. Homologs of unknown ORFs were identified using BLASTP (WU-BLAST2). The lists of genes having amino-acid sequence similarity or interaction with each unknown ORF were also analysed using GO Term Finder and common terms were identifed for the members of each list. The significance of each term was calculated and is presented here as further evidence in the process of specification. The terms assigned to ORFs using AVID (Jing and Keatling, 2005), GeneFAS (Joshi *et al.,* 2004) algoritms and functional linkages found by STRING (von Mering *et al.,* 2005) were also included for comparison when available (Table B.7).

6.6. Unknown ORFs Upregulated at Nitrogen Limitation

In at least two of the experiments considered, transcription of YMR090w was up-regulated under nitrogen limitation when compared to glucose limitation and in our data set its expression was correlated with genes from glutamate synthesis process ($p < 10^{-5}$). It was proposed by Erasmus *et al*. (2003) that succinate synthesis from glutamate pathway is active at high sugar stress, which agrees with present data where genes from glutamate synthesis process are up-regulated by the high glucose concentration in nitrogen-limited media. YMR090w has sequence similarity with a selenomethionyl gene product and other oxidoreductases. In addition, its product has aspartate-semialdehyde dehydrogenase and dihydrodipicolinate reductase motifs (Dolinski *et al.,* 2005). The product of YMR090w physically interacts with Mae1p which catalyzes the oxidative decarboxylation of malate to pyruvate and with Atp14p which is subunit h of the F_0 component of the mitochondrial F_1F_0 ATP synthase (Ito *et al.,* 2001). The YMR090w transcript was up-regulated by genetoxicity in correlation with some aldo/keto reductases encoding genes, transport and energy-related genes (Caba *et al.,* 2005). Thus, YMR090w may encode a protein that catalyses a reaction in which ATP is involved and which is on the pathway from ammonia to glutamate synthesis or subsequent succinate synthesis involving metabolism of 4-aminobutanoate and succinate-semialdehyde.

YGL157w was up-regulated under nitrogen-limited conditions in more than one experiment under investigation. Our analysis showed that YGL157w is expressed in correlation with genes annotated to transport activity ($p < 0.001$). By LC-MS/MS, serine hydrolase FFFs (fuzzy functional forms, structural motifs for identification of functional sites) were searched in yeast proteome and UDP-galactose-4-epimerase, carbonyl reductase, estradiol-17-dehydrogenase and sepiapterin reductase catalytic sites were found in product of YGL157w (Baxter *et al.,* 2004). It was reported to be a target of Rgt1p (Kaniak *et al.,* 2004), which is involved in the glucose-signalling pathway. However, YGL157w and the ORFs having amino-acid sequence similarity to YGL157w (YDR541c, YGL039w and YOL151w) are annotated to molecular function term dihydroflavanol 4-reductase due to their structural homology (Delneri *et al.,* 2000) and to oxidoreductase activity by direct assay (Katz *et al.,* 2003). Its product is found to reduce ethyl acetoacetate (Katz *et al.,* 2003b).

6.7. Unknown ORFs Upregulated at Glucose Limitation

Expression of the ORFs discussed below was up-regulated under glucose limitation when compared to nitrogen limitation in at least two of the experiments considered in this study. The up-regulation may be due to respiratory activity, the absence of glucose repression, or other cellular activities taking place under glucose limitation but not under nitrogen limitation.

Our analysis suggests that the YLR327c product may be involved in glycogen and energy reserve metabolism ($p < 10^{-10}$) or may have transferase function ($p < 10^{-4}$). YLR327c has amino-acid sequence similarity to Stf2p (YGR008c), a protein involved in regulation of the mitochondrial F_1F_0-ATP synthase. In the present data set, YLR327c and *STF2* expression are found to be correlated ($\rho = 0.97$). YLR327c is adjacent to *NMA1* (*YLR328c*) and *STF2* is neighbor to *NMA2* (YGR010w – amino-acid sequence similarity with *NMA1*), which encode nicotinic acid mononucleotide adenylyltransferases. YLR327c was reported to be down-regulated by cytotoxic stress (ethanol and NaCl) and up-regulated by genotoxic stress (Caba, 2005) in coordination with *STF2*. Moreover, it is reported to be a putative protein with cytoplasmic ribosome function and its functional category is

reported to be assembly of protein complexes (Güldener *et al.,* 2005). It also interacts with Los1p (Ito *et al*., 2001) which interacts with proteins involved in protein folding and organelle organization/biogenesis. The YLR327c product interacts with Elp6p, (Ito *et al.,* 2001) which is a part of the elongator of the HAP (histone acetyltransferase protein) sub-complex. Elp6p is required for elongator structural integrity and histone acetyltransferase activity. It shares amino-acid sequence similarity with proteins responding to external stimulus (3/8paralogs, $p = 0.002$). Thus, YLR327c product might be involved in energy reserve or energy derivation metabolisms with a transferase function.

In the present data set, YGR243w expression is in correlation with genes active in energy derivation from oxidation of organic compounds and carbohydrate metabolism ($p < 10^{-6}$). YGR243w encodes a mitochondrial protein and it is located close to *ENO1* (*YGR254w*) on the genome, which may indicate coordinated expression of this gene with YGR243w. *ENO1*p is involved in glycolysis and is reported to be repressed by glucose (Dolinski *et al.,* 2005). Our analysis yields significant energy derivation and carbohydrate metabolism terms. However, the detailed literature survey conducted does not indicate further information in support of this annotation.

Our transcript data on YGR067c indicate a role in generation of precursor metabolites/energy ($p < 10^{-6}$). YGR067c has amino-acid sequence similarity with zinc finger and ADR1p DNA-binding domains (Bohm *et al.,* 1997). Adr1p is a carbon source-responsive zinc-finger transcription factor, required for transcription of the glucose-repressed gene *ADH2*, of peroxisomal protein genes, and of genes required for ethanol, glycerol, and fatty acid utilization (Young *et al., 2003)*. YGR067c was reported to be *CAT8*-dependent and has CSRE (carbon source responsive) site in its upstream region (Haurie *et al.,* 2001). 17/28 ORFs having amino-acid sequence similarity to YGR067c protein are involved in regulation of transcription and DNA-binding activities ($p<10^{-12}$). YGR067c also has interactions (Ho *et al.,* 2002; Hazbun *et al.,* 2002) with proteins involved in organelle organization and biogenesis ($p < 0.01$). It was found in a complex (Ho *et al.,* 2002) with proteins from polysomes ($p < 10^{-4}$). Thus YGR067c may encode a protein involved in transcriptional regulation, specifically in the transcription of genes related to energy metabolism.

YER121w is expressed in correlation with genes involved in the generation of precursor metabolites/energy ($p < 10^{-6}$) and phosphatase regulation ($p < 0.001$). The YER121w product interacts with Grr1p (Hazbun *et al.,* 2003), which is involved in carbon catabolite repression, glucose-dependent divalent cation transport, high-affinity glucose transport, and morphogenesis (Güldener *et al.,* 2005). No other evidence exists to support its relation to any biological process or molecular function.

Our transcription data show that expression of YGR236c (*SPG1*) is correlated with genes responding to water deprivation ($p < 10^{-4}$). YGR236c (*SPG1*) encodes a mitochondrial protein (Sickmann *et al.,* 2003) which is in interaction with the transcription factor Hap5p (Ito *et al.,* 2001). Its transcription is induced in aerobic conditions (ter Linde *et al.,* 1999) and it is required for survival at high temperatures (Martinez *et al.,* 2004). Its transcription is down-regulated in response to diammonium sulphate (DAP), together with genes involved in protein and nitrogen compound catabolism ($p < 10^{-4}$) (Marks *et al.,* 2003) and up-regulated more than two fold at high sugar stress (Erasmus *et al.,* 2003) with other genes from carbohydrate metabolism ($p < 10^{-6}$). These findings do not exclude the function assigned by this study.

The present results indicate YFR017c to be involved in generation of precursor metabolites/energy pathways and to have aldo-keto reductase activity ($p < 0.001$). It is a member of a protein complex (Ho *et al.,* 2002) which concerns genes functional in energy reservation metabolism ($p<10^{-6}$) and which is located at bud neck ($p<10^{-4}$). The YFR017c product has interactions with proteins (Ho *et al.,* 2002; Hazbun *et al.,* 2003; Ito *et al.,* 2001) which are involved in glycogen/glucan metabolism or energy reserve/generation metabolism ($p < 0.005$). This evidence supports its possible involvement in energy-reserve pathways.

YER067w was expressed in correlation with genes acting on energy-generation metabolism ($p < 10^{-4}$) in the present data set. It is strongly downregulated by cytotoxic stress (ethanol and NaCl) (Caba *et al.,* 2005). Its product was found in a complex (Ho *et al.,* 2002) with ribosomal proteins ($p < 0.005$). Its expression responds to ergosterol, which is reported to have an important role in mitochondrial respiration (Bammert *et al.,* 2000). It is also reported to be the only gene which is regulated by both ploidy and the diauxic shift

(Galitski *et al.,* 1999). Its product was located in cytoplasm and nucleus by direct assay (Huh *et al.,* 2003), also the products of its paralogs were also located in the nucleus (5/6, $p < 0.01$). None of the available information excludes the possibility that YER067w may be involved in energy derivation.

The mitochondrial protein (Sickmann *et al.,* 2003) encoded by YKL187c is proposed to have a function in carbohydrate metabolism by our current analysis ($p < 10^{-4}$). Its paralogs respond to stimuli involved in cell communication and signaling ($p < 10^{-5}$). The protein products of some of its paralogs are located in the cell wall and internal membranes ($p < 10^{-5}$), functioning as transmembrane receptors ($p < 10^{-7}$). It expression is induced by glycerol (Richard *et al.,* 1997). Its product was reported to be in the fatty acid elongation subfamily of non-transporting membrane proteins (de Hertogh *et al.,* 2002). Further, its structural homologs are internalin proteins from *Listeria monocytogenes,* which are virulence proteins that enable cell-to-cell adhesion (Dolinski *et al.,* 2005). YKL187c expression is induced by diauxic shift and in the absence of the Srb10 complex, which disappears during nutritional stresses along with the products of other genes involved in the nutrient stress response and in the morphological change that permits searching for nutrients (Holstege *et al.,* 1998). Thus, Ykl187p might be located in mitochondrial membrane with a receptor function related to carbohydrate metabolism.

YMR107w (*SPG4*) expression is correlated with that of genes involved in intra-golgi transport and v-SNARE (vesicle membrane receptor) activity ($p < 10^{-3}$). The proteins with amino-acid sequence similarity to YMR107w are involved in organelle organization/biogenesis ($p < 0.1$). It is an essential gene for survival at high temperatures (Martinez *et al.,* 2004). Its transcription is induced under aerobic conditions (Piper *et al.,* 2002, ter Linde *et al.,* 1999). The TOR kinase complex controls cell growth by regulating translation, transcription, ribosome biogenesis, nutrient transport and autophagy in response to nutrient availability (Martin and Hall, 2005; Crespo and Hall, 2002). It has been reported that the YMR107w product has negative role in TOR signalling, which is upregulated by rapamycin and ymr107w deletants are rapamycin resistant (Xie *et al.,* 2005). The transcription of YMR107w is induced by diauxic shift and is negatively regulated by the Srb10 complex with other genes involved in the nutrient stress response and in the morphological change that permits foraging for nutrients (Holstege *et al.,* 1998).

YMR107w probably responds to nutritional stresses. However, this finding does not exclude the function suggested by this study, it may be involved in transmitting nutritional stress signals to Golgi.

YMR206w expression is correlated with that of *CAT8* and *PIG1* which are members of a cluster if genes involved in the regulation of carbohydrate biosynthesis (2/19, $p < 10^{-3}$) in present data set. No information is available in the literature relating to this ORF.

6.5. Unknown ORFs Down-regulated in a *hap4Δ/hap4Δ* deletion strain

YJL103c expression was found to be significantly downregulated in *hap4Δ/hap4Δ* mutant in both available data sets (Pir *et al.*, 2004 and Buschlen *et al.*, 2003). In the present data set, its expression was very strongly correlated with genes involved in oxidative phosphorylation ($p < 10^{-13}$). 7/25 of its paralogs are positive regulators of DNA-dependent transcription ($p < 10^{-9}$). 18 of these paralogous proteins were found in nucleus ($p < 10^{-5}$). It is reported that it is involved in transcriptional control, and its predicted product has a Cys6 cysteine - zinc cluster domain. Also, it is a putative heme-dependent regulatory protein and it has a cytochrome binding site (Güldener *et al.,* 2005). The Yjl103c protein interacts (Hazbun *et al.,* 2003) with the products of genes involved in nuclear positioning ($p<0.005$). The product of YJL103c was recently annotated to be involved in oxidative phosphorylation, as inferred by sequence and structure similarity (Deng *et al.,* 2005), and was defined as a zinc cluster protein by phenotype screening (Akache *et al.,* 2001). Supported by the above evidence, it can be concluded that YJL103c should be encoding a protein that is involved in transcriptional regulation of the respiratory pathway of yeast.

6.9. Unknown ORFs Up-regulated at High Growth Rate

The expression of YGR173w *(RBG2)* is correlated with that of genes from purine salvage, tRNA modification and ribonucleoside monophosphate biosynthesis pathways ($p<10^{-4}$). Its paralogs are involved in ribosome-nucleus export and functions in GTP-binding ($p<10^{-3}$ and 10^{-7}, respectively). Its product has a GTP1/OBG (GTP binding family) signature. It was stated that two cytosolic Obg proteins Rbg1p and Rbg2p (which share 52% amino-acid sequence identity) associate with translating ribosomes and it is possible

that they are involved in stress response (Datta *et al*., 2005). The YGR173w product is assigned to biological process GTP binding by amino-acid sequence similarity (Li *et al*., 2000). The available information does not include any supporting evidence for the present results.

YJL200c expression was also up-regulated at higher growth rates and is correlated with genes involved in transcription from RNA polymerase I promoter which promotes the transcription of rRNA genes for the precursor of the 28S, 18S, and 5.8S molecules ($p < 0.0005$). It was annotated with the term "aconitate hydratase activity" by sequence homology (Tatusov *et al*., 2000) and the location of this protein in the mitochondrion has been determined by direct assay (Huh *et al*., 2003; Sickmann *et al*., 2003; Ohlmeier *et al*., 2004).This protein was reported to be an acotinase, relying on the results of MALDI-TOF, 2-D gel analysis and amino-acid sequence identity (Shevchenko *et al*., 1996). It has motifs of aconitases and structural homology with aconitases from many species including bovine, pig and *E.coli* (Dolinski *et al*., 2005). It is reported to be involved in the TCA cycle (Przybyla-Zawislak *et al*., 1999; Dolinski *et al*., 2005) and the serine-isocitrate lyase pathway, catalyzing the conversion of citrate to cis-aconitate (Dolinski *et al*., 2005). The transcription of YJL200c was investigated on glucose and ethanol and found to be enhanced on glucose and decreased on ethanol (van der Berg *et al*., 1998). Although YJL200c was referred to as *ACO2* (as an isoform of *ACO1*/*GLU1,* which encodes aconitase*)* by Przybyla-Zawislak (1999) and van den Berg (1998), the name *ACO2* was not mentioned in recent publications and databases. Its paralogs are involved in amino-acid biosynthesis ($p < 10^{-4}$) and hydro-lyase activity ($p < 10^{-8}$). Response of YJL200c expression to amino acid availability revealed that it may be involved in the lysine biosynthetic pathway (Giaever *et al*., 2002). Its expression is down-regulated in response to the addition of lithium to galactose, together with that of with genes involved in transcription, translation, and nucleotide metabolism (Bro *et al*., 2003). Also its expression is down-regulated in strains deleted for *CHD1*, a gene thought to have a function in transcription and chromatin remodelling. (Tran *et al*., 2000). Other genes down-regulated in *chd1* deletants are involved in amino-acid and organic acid metabolism ($p < 10^{-7}$). Moreover, Gcn4p*,* which is involved in transcriptional activation of amino-acid biosynthetic genes in response to amino-acid starvation, binds to to the YJL200c promoter (Yeang *et al*., 2004). YCL200c transcription is downregulated more than two-fold under

high sugar stress (Erasmus *et al*., 2003), along with other genes from nucleotide, amino-acid and organic acid metabolism ($p < 10^{-8}$). Furthermore, it is up-regulated in response to DAP (Marks *et al*., 2003), together with genes involved in amino acid, nitrogen and organic acid metabolism ($p < 10^{-35}$). Present results and the evidence cited above indicate that YCL200c may be involved in transcriptional regulation of amino-acid biosynthesis pathways.

YNL175w (NOP13) expression was up-regulated at higher growth rates in both the present data set and in that of Grotkjaer (2005). Both of the genes (*LOC1* and *IMP4*) with correlated expression to YNL175w are involved in ribosome biogenesis ($p < 0.001$). Its transcript is down-regulated more than two-fold under high sugar stress (Erasmus *et al*., 2003), along with genes from nucleotide, amino acid and organic acid metabolism ($p < 10^{-8}$). The YNL175 protein interacts with proteins (Ho *et al.,* 2002; Hazbun *et al*., 2003; Ito *et al.,* 2001) involved in ribosome biogenesis ($p < 10^{-7}$) and was found in three complexes (Ho *et al*., 2002) with proteins involved in ribosome biogenesis ($p < 10^{-8}$). Ynl175p was also found in another complex with proteins acting in chromatid segregation and localized in the nuclear cohesion complex ($p < 10^{-8}$). It has amino-acid sequence similarity with genes from RNA metabolism ($p < 10^{-12}$). YNL175w was assigned the function RNA-binding by sequence homology and its location was determined to be nucleus and nucleolous by direct assay (Wu *et al*., 2001). Interacting gene products, proteins with amino-acid sequence similarity and proteins from the same complexes are also located in the nucleus, nucleolous and non-membrane-bound organelles, with high significance ($p < 10^{-8}$). When all this evidence is taken into consideration, YNL175w can be concluded to have a function in ribosome biogenesis and RNA metabolism.

6.10. Unknown ORFs Down-regulated at High Growth Rate

Expression of YMR090w is down-regulated at high growth rates, has been discussed already as a gene up-regulated under nitrogen limitation; we have annotated it as a gene acting on the glutamate biosynthesis pathway.

YOR173w (*DCS2*) expression is correlated to genes from carbohydrate and trehalose metabolisms in present data set ($p < 10^{-4}$). *DCS2* and its paralog, the *DCS1* gene, are

induced by heat shock and mutagens and are repressed by cAMP withdrawal (Kwasnicka *et al*., 2003). The YOR173w product interacts (Ho *et al*., 2002; Gavin *et al*., 2002; Hazbun *et al*., 2003; Ito *et al*., 2001) with genes involved in mitosis and enzyme regulation ($p < 0.005$). It was found in a complex (Gavin *et al.,* 2002) with proteins located in the transcription factor TFIIH complex ($p<10^{-10}$), and in another complex (Ho *et al.,* 2002) with members involved in regulation of redox homeostasis that has thioredoxin activity ($p<0.0002$). The AVID algorithm (Jiang and Keatling, 2005) predicted that it is involved in deadenylylation-dependent decapping and ubiquitin-specific protease activity. It is reported to be regulated by Msn2p and Msn4p, two transcriptional activators that mediate the response to stress conditions, and the Ras-cAMP-cAPK signalling pathway, which is essential for the control and integration of regulation of growth, cell cycle progression and metabolic activities, in particular with respect to the nutritional status (Pedruzzi *et al.* 2000). YOR173c product has the structure of dcps (mRNA decapping enzyme) and was reported to be a scavenger mRNA decapping enzyme. It was described as a non-essential protein containing a HIT (histidine triad) motif (Dolinski *et al.,* 2005). YOR173c is up-regulated in a strain with *RNS4* gene disrupted, along with other genes responding to stress and involved in catabolism ($p < 0.005$) (Unno *et al.,* 2005) . Its expression is negatively regulated by cAMP, along with other genes involved in carbohydrate metabolism ($p < 10^{-6}$) (Tadi *et al.,* 1999). Tadi *et al.* (1999) demonstrated a great accumulation of its transcript when glucose becomes limiting, which suggests that it encodes a product, required in larger quantities, that functions in the physiological transitions due to nutritional limitations. Thus, present results are partly supported by the available evidence.

The transcription of YDR070c (*FMP16*) is correlated with that of genes responding to water, according to the present data set ($p < 0.0005$). Its product is localized in the mitochondrion (Huh *et al*., 2003; Sickmann *et al*., 2003) and it is reported to be induced under aerobic conditions (ter Linde *et al*., 1999). It is also induced by the diauxic shift and is negatively regulated by the Srb10 complex, along with other genes involved in the nutrient stress response and in the morphological change that permits foraging for nutrients (Holstege *et al.,* 1998). YDR070c is upregulated by genotoxicity, down-regulated by toxicity, together with a sub-set of transport and energy genes (Caba *et al*., 2005). Its transcription is down-regulated in response to the addition diammonium sulphate (DAP), together with genes involved in protein and nitrogen-compound catabolism ($p < 10^{-4}$)

(Marks *et al.,* 2003). YDR070c expression is up-regulated more than two-fold upon high-sugar stress (Erasmus *et al.,* 2003), along with other genes from the carbohydrate metabolism category ($p < 10^{-6}$). It was reported to be up-regulated in response to Bax, a regulator of apoptosis, together with genes acting on glucose and hexose metabolism ($p < 0.0005$) (Reekmans *et al*., 2005). It is also up-regulated in absence of *SAN1*, a gene encoding ubiquitin-protein ligase (Gardner *et al*., 2005),along with other genes overrepresented in GO biological process terms stress response, and protein folding and molecular function term oxidoreductase activiy ($p < 10^{-4}$). It was induced by Pdr3p and Pdr1p, transcriptional activators of pleiotropic drug resistance (Onda *et al*., 2004), and other genes responsive to stress ($p < 0.005$). Thus, present results are partly supported by the available evidence and suggest that *YDR070c* is a stress-responsive gene, probably responding to nutritional stresses.

The annotations proposed in this study and existence/absence of external evidence are given in Table 6.4.

6.11. Precision of Present Annotations

Correlation of gene expression is assumed to be an indication of shared functions in this study. Validation of this assumption and evaluation of the precision of the annotations are made by re-annotating the known genes.

Genes whose expression changed significantly in at least two of the experiments were analysed for validation purposes, and the results are plotted in Figure 6.1. An annotation is considered as "correct" only if the term allocated in this study is exactly the same as the known term for the gene. Thus, the criterion of correctness is stringent. Depending on the threshold p-value used, 15 to 45 of the total of 177 genes included in the analysis can be annotated by at least one correct term. If the threshold p-value is below $1x10^{-6}$, the fraction of correctly annotated genes is above 0.70 and the fraction of correctly assigned terms is above 0.30. However, for $p > 0.001$, the fractions above remain below 0.2 and 0.1 respectively.

Table 6.4. Annotations proposed in this study

ORF	Proposed function or process	External Evidence
YJL103c	Transcriptional regulation of respiration ($p < 10^{-13}$)	Yes
YLR327c	Glycogen and energy reserve metabolism ($p < 10^{-10}$), transferase function ($p < 10^{-4}$)	Yes
YGR243w	Energy derivation and carbohydrate metabolism ($p < 10^{-6}$)	Yes
YGR067c	Generation of precursor metabolites and energy ($p < 10^{-6}$)	Weak
YMR090w	Glutamate synthesis pathway ($p < 10^{-5}$)	Yes
YOR173W (DCS2)	Carbohydrate and trehalose metabolism ($p < 10^{-4}$)	Weak
YKL187c	Carbohydrate and alcohol metabolism ($p < 10^{-4}$)	No
YGR236c (SPG1)	Water deprivation ($p < 10^{-4}$)	No
YER067w	Energy derivation metabolism ($p < 10^{-4}$)	No
YGR173W	Purine salvage, tRNA modification and ribonucleoside monophosphate biosynthesis pathways ($p < 10^{-4}$)	No
YFR017c	Generation of precursor metabolites and energy ($p < 10^{-3}$)	Yes
YER121w	Generation of precursor metabolites and energy ($p < 10^{-6}$), phosphatase regulation ($p < 10^{-3}$).	Weak
YDR070C (FMP16)	Response to water ($p < 10^{-3}$)	Weak
YGL157w	Transport and carboxylic ester hydrolase ($p < 10^{-3}$)	No
YMR107w (SPG4)	Intra-golgi transport, v-SNARE activity ($p < 10^{-3}$)	No
YMR206w	Regulation of carbohydrate biosynthesis ($p < 10^{-3}$)	No
YNL175C (NOP13)	Ribosome biogenesis and RNA metabolism ($p < 10^{-3}$)	Yes
YJL200C	Transcription from RNA polymerase I promoter ($p < 10^{-3}$)	No

When all genes are included in the re-annotation analysis without any filtering process, the fraction of correctly annotated genes is reduced to about 0.45 while the fraction of correctly assigned terms goes down to about 0.25; p-value below $1x10^{-6}$. For $p > 0.001$, the fractions remain unchanged, *i.e.* below 0.2 and 0.1, as in the previous case (results not shown).

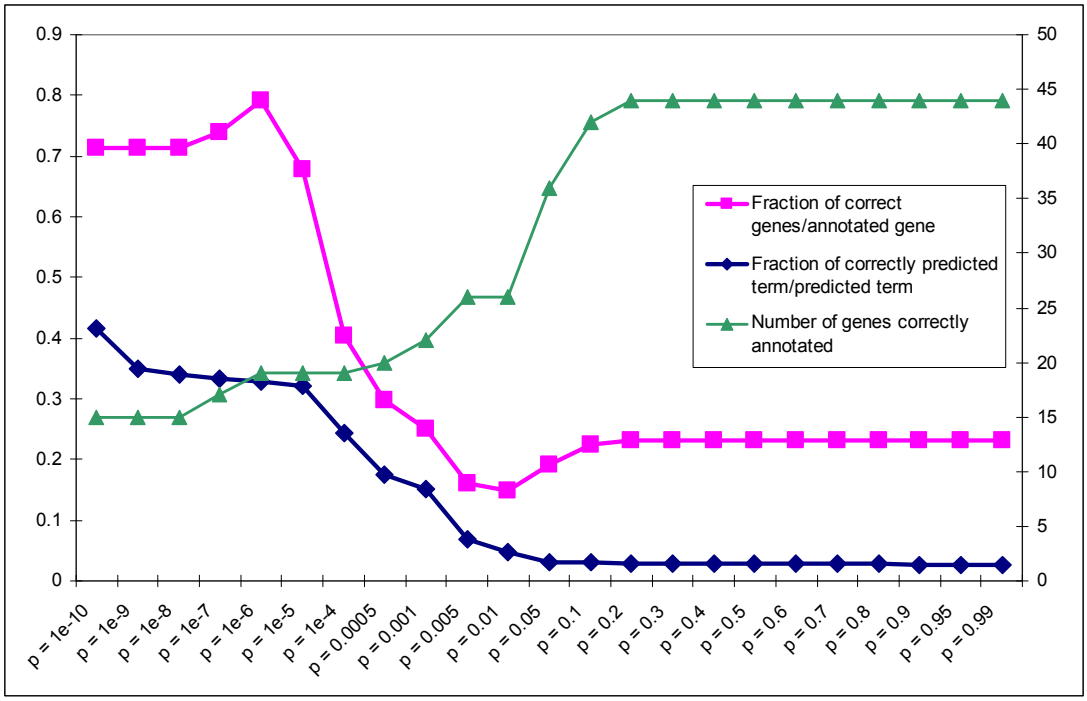

Figure 6.1. Number of correctly annotated genes and fractions of correct annotations.

Thus, the precision of the annotations made is improved by using linear modeling and incorporating the results of multiple experiments for filtering out genes that are not differentially expressed under the selected conditions, although the number of genes that can be annotated is considerably reduced by this approach.

The annotations made in this study are generally based on p-value thresholds that are in the low-precision region of the curve, and they should be considered as "validated" only if external evidence exists. However, the annotations of the terms with low p-values ($p < 10^{-5}$) have relatively higher precision, *i.e.* the fraction of correctly annotated genes is about 0.65 and the fraction of correctly assigned terms is about 0.30.

6.9. Comparison of Present Results with Automated Annotation Results

The annotations generated by automated algorithms for the ORFs investigated here are given in Table B.7. GeneFAS (Joshi *et al.,* 2004) generated many results for all of the ORFs, accompanied by a probability and a significance value. The results with a probability higher than 0.25 are included in the following comparison. GeneFAS produces

general terms like “cell growth”, for instance, “cell growth and/or maintenance” is the top scoring annotation for all ORFs.

AVID produces biological process results for 10/17 genes under investigation and STRING (von Mering *et al*., 2005) provides functional linkages for 14 genes. The functional linkages provided by STRING cover only part of the information that actually exists in literature and databases. Text mining in Abstracts of the publications may sometimes lead to ambiguous information.

For YMR090w, the STRING linkage is to Mae1p (NAD-dependent malic enzyme), which is a connection used in the present analysis as well. AVID produces no predictions, and the 17 predictions of GeneFAS with p >0.25 (cell growth and/or maintenance, metabolism, transport, cell communication, response to external stimulus, response to stress, cell organization and biogenesis, catabolism, homeostasis, protein metabolism, amine metabolism, amino acid and derivative metabolism, organic acid metabolism, nucleobase, nucleoside, nucleotide and nucleic acid metabolism, response to biotic stimuli, oxygen and reactive oxygen species metabolism, response to oxidative stress) is not related to the glutamate synthesis pathway.

For YGL157w, the automated annotations disagree with the transport and carboxylic ester hydrolase functions reported here. Some of the literature information presented here is reflected by linkages found by STRING. STRING links YGL157w with two putative reductases, a putative glycerol dehydrogenase and a glucose-6-phosphate isomerase. AVID annotates fatty acid and galactose metabolism. GeneFAS results in cell growth and/or maintenance, metabolism, nucleobase, nucleoside, nucleotide and nucleic acid metabolisms with probability greater than 0.25.

AVID makes no predictions for YGR243w. The STRING links this ORF with a succinyl-CoA ligase and a zinc-finger protein. GeneFAS results in cell growth and/or maintenance, metabolism, cell communication and signal transduction with probability greater than 0.25. These annotations may be related to present annotation “energy generation and carbohydrate pathways”.

YLR327c and YER121w, although annotated using evidence from this study, were not annotated by AVID or STRING. GeneFAS annotated the terms cell growth and/or maintenance, metabolism, response to stress, cell organization and biogenesis, cell communication, cell adhesion, response to external stimulus, nucleobase, nucleoside, nucleotide and nucleic acid metabolism and carbohydrate metabolism for YLR327c and cell growth and/or maintenance, cell organization and biogenesis, metabolism, and transport for YER121w. These terms are not directly related to the "transferase function, glycogen and energy reserve metabolism" terms annotated to YLR327c and "phosphatase regulation, generation of precursor metabolites and energy" terms annotated to YER121w in the present study.

YFR017c was annotated to trehalose metabolism by AVID and linked to proteins involved in pyridoxine biosynthesis, maintenance of ploidy, ubiquitin carboxyl-terminal hydrolase, and a protein that binds proteins and DNA by STRING which may be related to the "generation of precursor metabolites and energy" annotation made here. GeneFAS annotated this term in 9th place among 27 terms with probability greater than 0.25.

The "generation of precursor metabolites and energy metabolism" annotations proposed for YGR067c are similar to the 5th-ranked prediction made by GeneFAS (energy pathways) among 20 terms with probability greater than 0.25. The predictions of AVID (G1-specific transcription in mitotic cell cycle and metabolism) and links of STRING (proteasome component, 26S proteasome regulatory subunit, essential RNA-binding G-protein effector of mating response pathway) are irrelavant to the present results.

YKL187c is linked to a probable succinate dehydrogenase by STRING and annotated to 15 terms (probability greater than 0.25) by GeneFAS (cell growth and/or maintenance, metabolism, organic acid metabolism, energy pathways, transport, electron transport, phosphorus metabolism, cell communication, carboxylic acid metabolism, carbohydrate metabolism, energy derivation by oxidation of organic compounds, coenzymes and prosthetic group metabolism, signal transduction, oxidative phosphorylation, phosphate metabolism, response to stress, main pathways of carbohydrate metabolism, phosphorylation biosynthesis, TCA intermediate metabolism, main pathways of carbohydrate metabolism, amine metabolism, amino acid and derivative metabolism,

tricarboxylic acid cycle, and lipid metabolism), most of which are related to the un-supported "carbohydrate and alcohol metabolism" annotation made in this study. AVID made no predictions for this ORF.

The present annotations made for YGR236c, YMR107w, YMR206w, and YER067w are not supported by external evidence. YGR236c is linked, by STRING, to proteins with the descriptions transcriptional activator, bud-site selection, glucose-6-phosphate isomerase. The GeneFAS annotations for YGR236c are cell growth and/or maintenance, metabolism, cell organization and biogenesis, energy pathways, organic acid metabolism, budding, cytoplasm organization and biogenesis, carbohydrate metabolism, vitamin metabolism, lipid metabolism, energy derivation by oxidation of organic compounds (probability greater than 0.25). These terms are not related to present annotation "water deprivation". No linkages were found by STRING for YMR107w, GeneFAS resulted in cell growth and/or maintenance, metabolism, transport and cell cycle (probability greater than 0.25) which are irrelevant to present annotations "intra-golgi transport, and v-SNARE activity" for YMR107w. STRING links YMR206w to a sirohaeme biosynthesis protein. Among 24 annotations made with probability greater than 0.25 by GeneFAS, "carbohydrate metabolism" is in 6^{th} rank, which is similar to the present annotation "regulation of carbohydrate biosynthesis" for YMR206w. STRING links YER067w to enzymes acting on sterols, ribosomal proteins and a cytochrome. GeneFAS annotation "energy pathways" is in 7^{th} rank among 14 annotations with probability greater than 0.25 matches the present annotation "energy derivation metabolism". AVID made no predictions for these ORFs, except for YER067w (regulation of cell redox homeostasis).

YJL103c was annotated to transcriptional regulation of respiration with strong evidence in this study, similar predictions were made by GeneFAS though followed by many irrelevant terms (oxidative phophorylation ranked 20th among 36 annotations). AVID made no prediction and all STRING linkages are irrelavant (probable serine/threonine-protein kinase, 26S protease subunit, aryl-alcohol dehydrogenase, aspartyl protease precursor, sporulation-specific mitogen-activated protein kinase)

YNL175c was annotated similar terms related to ribosome biogenesis and RNA metabolism by all tools, which is supported by the evidence of the present work.

YJL200c is predicted to be involved in transcription from RNA polymerase I promoters with no additional evidence in the present work. GeneFAS annotations are cell growth and/or maintenance, cell cycle and metabolism (probability greater than 0.25). STRING links YJL200c with isocitrate dehydrogenases, AVID annotates tricarboxylic acid cycle, citrate metabolism and propionate metabolism. The annotations/linkages by AVID and STRING are in accordance with some of the information available in literature.

YDR070c, was linked to a glucosamine 6-phosphate N-acetyltransferase and an activator of multidrug resistance genes by STRING. Fourth among the 16 annotations made by GeneFAS is “response to stress”, which is a generalized term when compared to the present annotation “response to water” which was not supported by any evidence. AVID made no annotation for YDR070c.

For YGR173w, weak evidence was found for the present annotation “purine salvage, tRNA modification and ribonucleoside monophosphate biosynthesis pathways” and GeneFAS annotated similar terms (cell growth and/or maintenance, metabolism, cell organization and biogenesis nucleobase, nucleoside, nucleotide and nucleic acid metabolism, cytoplasm organization and biogenesis, transport, ribosome biogenesis and assembly, transcription, RNA metabolism biosynthesis transcription, DNA-dependent ribosome biogenesis, RNA processing, organic acid metabolism, carbohydrate metabolism, aromatic compound metabolism, heterocycle metabolism, ion transport, catabolism). The AVID prediction is “from ER to Golgi transport” and STRING findings are “low affinity Fe(II) transport protein, invertase 2 precursor, and a protein with similarity to mammalian monocarboxylate permeases”. These results are unrelated to the present results.

The annotations “carbohydrate and trehalose metabolism” for YOR173w were supported by only weak evidence. AVID made the annotations “deadenylation-dependent decapping” and “metabolism”. GeneFAS predicts “metabolism” in second place among 34 annotations. STRING links YOR173w with a cyclin, a serine/threonine-protein kinase, a subunit of TFIIH and nucleotide excision repair factor 3 complexes, a scavenger mRNA decapping enzyme, a transcriptional activator, a eukaryotic translation initiation factor and an aminopeptitase precursor, none of which is related to present results.

6.11. Success of Proposed Method

In present data set and the other experiments under consideration, the genes whose expression levels vary significantly fall into the biological process categories: transport processes, carbohydrate metabolism, electron transport, cellular respiration, generation of precursor metabolites and energy central metabolism. This reflects the fact that the perturbations applied were directly related to growth rate, carbon and nitrogen metabolisms. Thus, the unknown ORFs with significant expression changes were strongly correlated with genes involved in pathways of central metabolism. Consequently, the success of the annotations is context-dependent, perturbations that affect other pathways would result in the annotation of a completely different set of unknown ORFs.

Our methodology involves linear modeling and subsequent functional analysis of genes and ORFs whose expression changes significantly. Annotations for 17 of the investigated ORFs are proposed and 7 of these annotations are supported by external evidence. For the 10 ORFs, the annotations made depend solely on transcriptome data supported by weak, or no, external evidence. It should be kept in mind that the predictions made with low-precision p-values should be considered only if they are supported with external evidence. All annotations with $p < 10^{-5}$ were supported with such evidence, and so this value may be considered as a safe cut-off limit for such annotations.

Re-annotation of known genes selected by linear modeling and combining the results from independent data sets demonstrated that 30% of the annotations made with high significance ($p < 10^{-6}$) are correct and 70% of the genes are annotated at least one correct term.

Comparison of present annotations to those produced by automatic annotation tools indicated that automated annotations may miss some of the information in the literature, which may give misleading results.

Since tedious investigations of all evidence are needed for the annotation step following a transcriptome experiment, we investigated only the most affected ORFs, and satisfactory results superior to those obtained with current automated algorithms were

obtained. Implementation of this methodology automatically is possible only when software with a rigorous decision algorithm is available. Next section will be focused on development of a novel automated annotation tool, PredPower.

7. INVESTIGATION OF PREDICTION POWER OF NETWORKS FOR FUNCTIONAL ANNOTATION

In functional genomics, besides direct experimental analysis methods, *in silico* methods making use of the assumption that "similarities among the properties of the genes may indicate shared functions" found wide-spread acceptance. In identification of functions of unknown genes, similarities in genomewide data on gene expression, gene product localization, amino-acid sequence similarity, transcription factor binding, phenotypes of deletion mutants and physical/genetic interactions among proteins/genes proved to be useful. Frequently, the methodology involves making estimations using the similarities among known properties of the genes and verifying the results using the existing classification information on genes. Finally, novel annotations to unknown genes are proposed on this basis.

Constructing biological networks helps the organization and analysis of available information on the properties of genes. The networks can also be used as tools for defining the extent of similarities between the genes. The genes close to each other on a network have a higher extent of similarity in the property under investigation, which may indicate a common function. Protein-protein or protein-DNA interaction networks, as well as transcriptional regulation networks, have frequently been used in functional genomics research. Networks depending on other properties of the genes can also be used for estimating functions of the unknown members of the network. In Section 2.13, a list of publications reporting methods for *in silico* functional genomics of *S. cerevisiae* and Gene Ontology is given (Table 2.4). This list is far from being complete, however it indicates the trends of computational functional genomics research.

The functions assigned to genes constitute large lists of terms or classes, thus these lists also need to be organized. The classification of functions and annotations are provided by various databases. The classifications and annotations provided by KEGG (Kanehisa and Goto, 2000), MIPS (Mewes *et al*., 2005) and SGD (Dolinski *et al.*, 2005) are the most frequently used ones for yeast *S. cerevisiae*. For many species, Gene Ontology Consortium provides Gene Ontology (GO) which is composed of detailed classes of "biological

processes", "molecular functions" and "cellular components" that aim to describe the gene products. These terms constitute an acyclic map of parent-child relations among general and specific classes of functions. For *S. cerevisiae*, SGD also provides the GO mapping of all genes and ORFs.

The methods for predicting the functions of genes and proposing new annotations need to be verified, also a "reliability index" should accompany each proposed annotation. Since reliability of the new annotations depends on the descriptive power of the biological information used, standardized evaluation of the "prediction power" of the biological networks on the classification networks is believed to be useful.

In this study, the aim was to provide a framework for investigating the prediction power of biological networks on classification networks and making new annotations for unknown members of the biological networks. The software developed, PredPower, can be used for any biological and classification network pair, provided that a set of relations (annotations) between the two networks is known.

7.1. Sources of Data and Network Construction

Pearson correlation coefficients among the ORFs were calculated using transcriptome data collected in present work. Amino-acid sequence similarity between the products of the ORFs were obtained using WU-BLAST2 (BLASTP) (Gish, 2004). Interactome data for *S. cerevisiae* were collected from three sources: GRID (Breitkrutz *et al.*, 2003), STRING (von Mering *et al.*, 2005) and BIND (Alfarano *et al.*, 2005) databases. Transcription regulation data were taken from "Yeast Search for Transcriptional Regulators And Consensus Tracking" database (YEASTRACT, 2005). Gene Ontology (GO) annotations and terms were downloaded from SGD (Dolinski *et al.*, 2005).

7.1.1. Construction of Networks

Two types of networks were constructed. Type 1 networks are the biological networks constructed using the similarity between the genes. The nodes are the genes in biological networks and similar genes are connected to each other by edges. The five

biological networks constructed are given in Table 7.1. The type 2 networks are classification networks. The nodes are the GO terms and edges are the parent-child relations between the terms. The two classification networks given in Table 7.2 were constructed for biological terms and molecular function terms.

7.2.3. Biological Networks

In the biological networks constructed, the nodes were the ORFs appearing in GO annotation lists of SGD. The edges connect the correlated ORFs (Pearson correlation coefficient thresholds of $\rho > 0.90$ and $\rho > 0.95$) in transcriptome networks.

In the amino-acid sequence similarity (homology) network, the edges connect the ORFs whose proteins are reported to have amino-acid sequence similarity by WU-BLAST2. BLASTP was run using sequence filtering with default values for each ORF. Only the top 100 homologs to the ORFs were selected when more than 100 homologs were found. In the interactome network, the ORFs reported to be interacting were connected by edges. In regulatory network, the ORFs regulated by at least two common TFs were connected by edges.

The compendium network was constructed by combining all the nodes and the edges in the five networks described, and the replicate edges were removed. Thus, even if a pair of nodes were connected in more than one network, only one edge connected them in the compendium network. All edges were considered to be identical in these networks, therefore neither “weight” nor “strength” scores were assigned.

In Table 7.1, the total number of nodes (ORFs) with at least one connection in each network and the total number of connections among the ORFs are given. The number of connected nodes in regulatory network is small when compared to other networks since only “verified” data from Yeastract are used and only the ORFs sharing at least two TFs are connected. However, the total number of connections is very large when compared to other networks.

Table 7.1. Dimensions of biological networks

Network	Nodes	Edges
Expression Correlation (r > 0.90)	6020	151944
Expression Correlation (r > 0.95)	4855	25610
Homology	6657	179223
Interactome	4778	74030
Regulation	2253	249073
Compendium	6702	631510

7.1.3. Classification Networks

The classification networks were composed of parent-child relations among the GO terms. The nodes were the terms and edges were the parent-child relations. Since GO terms constitute a directed acyclic graph, only top-to-down connections are included in the network. The relation among two networks were the annotations of the ORFs to the terms. Two classification networks were constructed for biological process and molecular function terms. Table 7.2 gives the number of nodes and edges in classification networks.

Table 7.2. Dimensions of classification networks

Network	Nodes	Edges
Biological Process (BP)	1841	16521
Molecular Function (MF)	1517	8321

7.1.4. Algorithm of PredPower

The terms assigned to neighboring nodes of each ORF in the biological network were listed using a classification network. Then the significance (p-value) of each term in the classification network for the ORF of interest was calculated based on the number of relations it has with the neighbors of the ORF. The p-values calculated constitute a matrix with the number of nodes of the biological network and number of nodes of the classification network as its dimensions. The significance of a term is higher, the higher the number of neighbors of the ORF annotated to the term; however, if none of the neighbors of the ORF is annotated to that term, p is equal to 1.0. New annotations with p-values

smaller than the threshold p-value are proposed among the ORFs and terms. 23 thresholds for p-values in the 1 x 10^{-10} to 0.99 range were tested for new relations, i.e., proposed annotations.

The prediction power of the biological network on the classification network was evaluated by comparing the existing annotations of the "known" ORFs with the annotations proposed. The total number of ORFs correctly annotated with at least one term, the fraction of the ORFs with at least one correct annotation and the fraction of the correctly annotated terms were calculated.

Some of the annotations in SGD are very specific and parents of those direct annotations are also descriptive. Thus, in addition to the analysis which considers only the direct annotations of the ORFs, analyses including the parents of the direct annotations were also made. The neighbors in the classification network included in the analysis were at the following levels: "no neighbors – direct annotations only", "up to first level neighbors", "up to second level neighbors ", "up to third level neighbors", "up to fourth level neighbors" and "all possible neighbors". These levels correspond to "direct annotations only", "direct annotations and the first level parents of them", etc., respectively. Table 7.3 gives the total number of annotations when selected levels of parents were included. A sample run is described in full detail for the networks given in Figure 7.1 in Appendix A.

Table 7.3. Dimensions of biological networks

Annotations	Biological Process	Molecular Function
Only direct annotations	11048	8267
Up to first level parents included	25932	17166
Up to second level parents included	41644	22439
Up to third level parents included	59000	26205
Up to fourth level parents included	76483	28822
All parents included	103681	32100

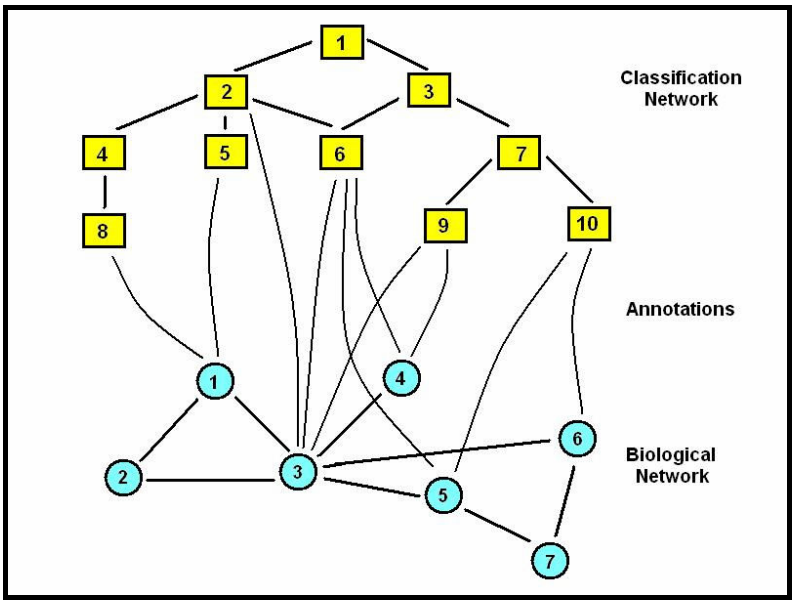

Figure 7.1. General structure of networks processed by PredPower.

7.3. Tests on PredPower

Prediction power of biological networks on GO prediction trees are evaluated considering the following three statistics:

a. Number of ORFs with at least one correctly predicted term

b. Ratio of all ORFs with a predicted term to ORFs with at least one correctly predicted term

c. Ratio of total number of predictions to number of correctly predicted terms.

Generally, the number of ORFs with at least one predicted term increases with the increasing p-values threshold and with increasing levels of neighbors included in the analyses. The fraction of correctly predicted ORFs is high at the stringent p-values, they decrease at intermediate p-value thresholds and increase again at very high p-values. Fraction of correctly made predictions decrease as the p-value threshold is increased.

The p-value threshold 1×10^{-7} is selected as an optimal point where both the fraction of correctly predicted ORFs and the fraction of correct predictions are high. However, there is a trade-off between high fractions and high number of predicted ORFs. The number of predicted ORFs remain low at the low p-values where both of the fractions

mentioned are high. Including all neighbors may improve both fractions of correct predictions and number of ORFs predicted, however, as the number of neighbors increased, newly included biological processes are more general (upper members of GO terms tree), which may be less descriptive. Investigation of all cases indicated that including 4 levels of neighborhood allows prediction of functions for a satisfactory number of ORFs while the most general terms are not included.

Many similar terms are predicted for each ORF in most of the cases. Frequently an ORF is connected with a set of genes that are involved in several closely related terms, consequently all of these terms are predicted for the ORF at similar significance levels. Thus, most of the terms predicted for an ORF are closely related to each other and indicate similar biological processes or molecular functions. However, only one or two of these terms are actually annotated to the ORF, which causes the overall fraction of correct predictions to be low.

All cases investigated are discussed in following sections. Comparison among the cases indicated that combining all biological networks in one large network enables higher prediction powers to be reached. 2981 biological processes and 1975 molecular functions for 443 and 524 unknown ORFs are predicted as an outcome of combining all biological information.

7.2.1. Prediction Power of Transcriptome Data on GO Classification Trees

Two correlation networks were constructed using the genes correlated to each other in present transcription dataset, the threshold values were 0.95 and 0.90. The prediction power of these networks were evaluated on GO biological process terms network.

The number of ORFs with at least one correctly predicted biological process increase as the p-value threshold for the predicted terms is increased (Figure 7.2A). Also increasing the level of neighbors included in the analysis, generally increases the number of ORFs correctly predicted. An exception is the 4 neighbor case, where the total number of correctly predicted ORFs exceeds the all-neighbors-included case for low significance predictions. The fraction of ORFs with at least one correctly predicted term to all predicted

ORFs decrease as the p-value threshold is increased up to 0.005. Beyond this threshold, the fraction increases (Figure 7.2B). However, the fraction of correctly predicted terms to all predicted terms generally decrease as the threshold is increased, and beyond p = 0.005, more than 90 per cent of the predictions made are incorrect (Figure 7.2C). The fractions generally increase as more neighbors are included in the analysis, the results obtained in 3 or 4 neighbors case approximate the results obtained in all-neighbors-included case. The highest fractions are observed for $p < 1 \times 10^{-10}$, about 60 per cent of the predicted ORFs that are predicted at least one correct term, while about 40 per cent of all predictions are correct.

In the present data set, 125 ORFs can be predicted using the parameters selected (4 neighbors, $p < 1 \times 10^{-7}$), and for 69 of the ORFs there is at least one term that is correctly predicted. 17 out of 125 ORFs are unknowns, and 89 novel predictions are made for these unknown ORFs. The expected fractions of correctly predicted ORFs and terms are 0.55 and 0.42 respectively for the parameters selected, which means that approximately 1 out of 2 ORFs predicted has at least one correctly predicted term and 1 out of 2 predictions is made correctly.

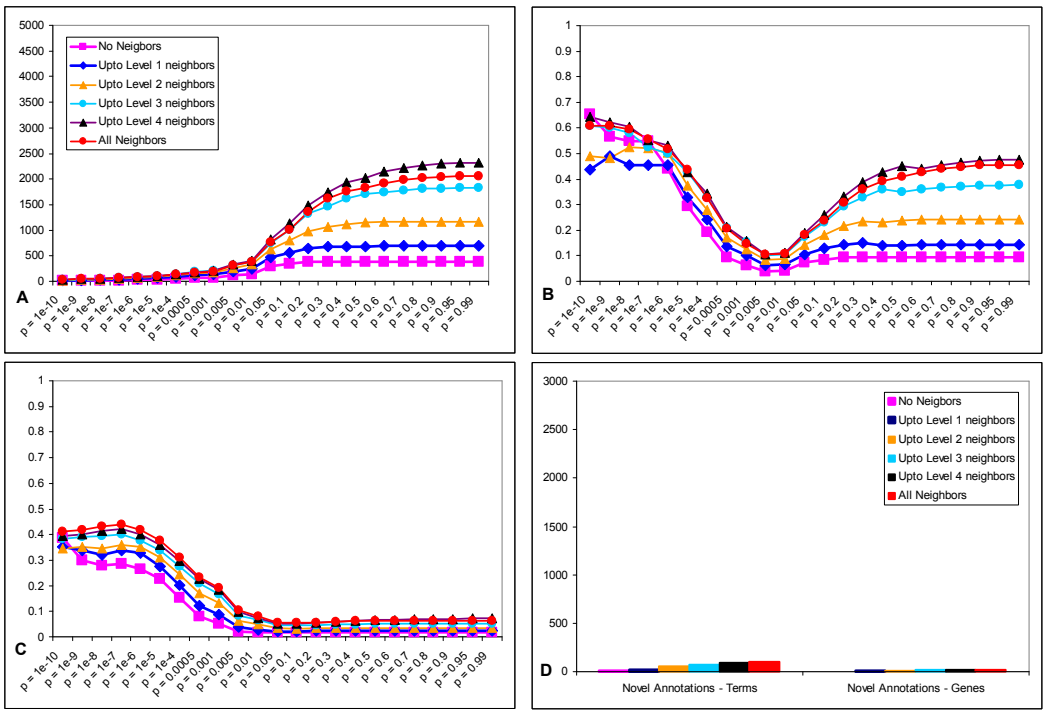

Figure 7.2. Expression correlation ($\rho > 0.95$), biological process terms

The limited number of ORFs predicted for given parameters can be regarded as an indication of low prediction power of this data set. The biological network constructed using the correlation coefficient threshold 0.90 results in similar prediction trends (Figure 7.3) but it allows prediction of more ORFs. Decreasing the correlation coefficient threshold causes the fraction of correct predictions to decrease when compared to $\rho > 0.95$ case (Figure 7.3B and C). The number of novel predictions increase in $\rho > 0.90$ case, using the parameters selected (4 neighbors, $p < 1 \times 10^{-7}$), 154 unknown ORFs are predicted, totalling 686 terms. At this significance level, 37 per cent of these ORFs are expected to be predicted at least one correct term, and 28 per cent of all predictions are expected to be correct.

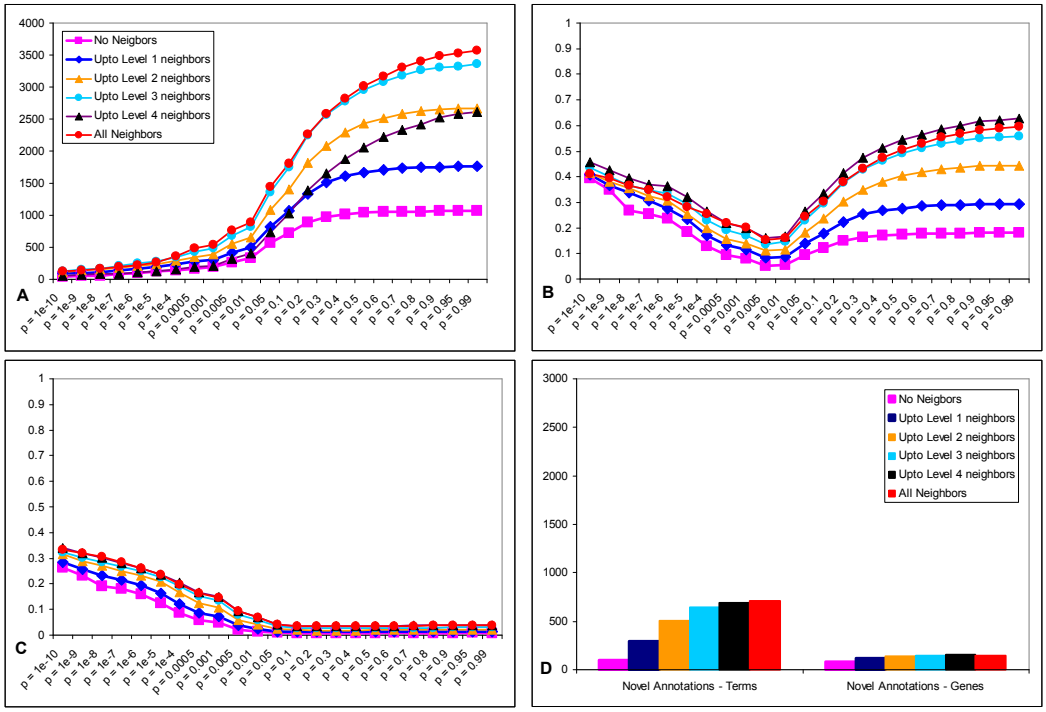

Figure 7.3. Expression correlation ($\rho > 0.90$), biological process terms

Prediction power of these two biological networks was also evaluated for network of molecular function terms. The number of ORFs predicted and fraction of correct predictions are given in Figures 7.4 and 7.5 for networks of correlation coefficients 0.95 and 0.90, respectively. The trends and fractions of correct predictions are similar; however, the number of predicted ORFs and number of novel predictions are lower when compared

to biological process classifications. This is an expected result of the lower number molecular function terms, accompanied by a lower number of predictions and fewer parent-child relations in the molecular function network as compared to the biological process network (Tables 7.3 and 7.4). 11 and 93 unknown ORFs are predicted to 44 and 289 molecular function terms in 0.95 and 0.90 networks correspondingly, when 4 neighbors are included and terms with $p < 1 \times 10^{-7}$ are taken into consideration. Expected fraction of correctly predicted ORFs and terms are 0.54 and 0.30 while fraction of correctly predicted terms are 0.42 and 0.26 in 0.95 and 0.90 cases, respectively.

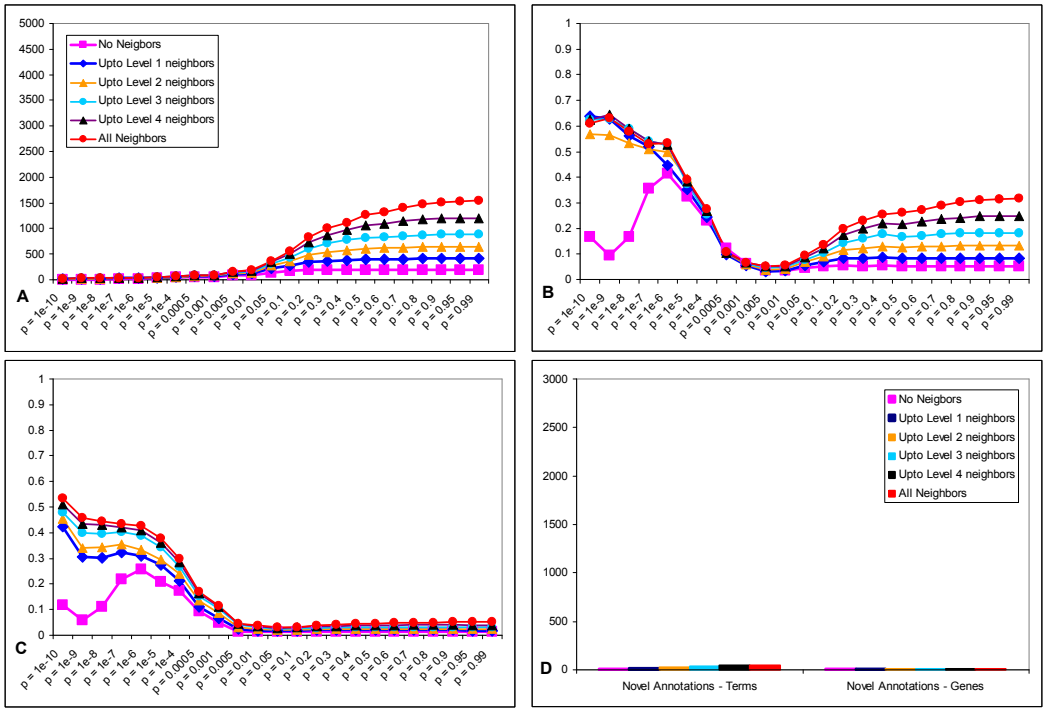

Figure 7.4. Expression correlation ($\rho > 0.95$), molecular function terms

7.3.2. Prediction Power of Protein Amino-acid Sequence Similarity on GO Classification Trees

The network of proteins having amino-acid sequence similarity was constructed using the results from WU-BLAST2 (BLASTP) analyses for 6657 gene products, all ORFs with homolog products were connected without any filtering, regardless of the extend of the amino-acid sequence similarity. The power prediction analysis was carried out using 23 levels of threshold p-values and levels of neighborhood to infer the prediction power of

interaction network on GO prediction tree of biological process and molecular function terms.

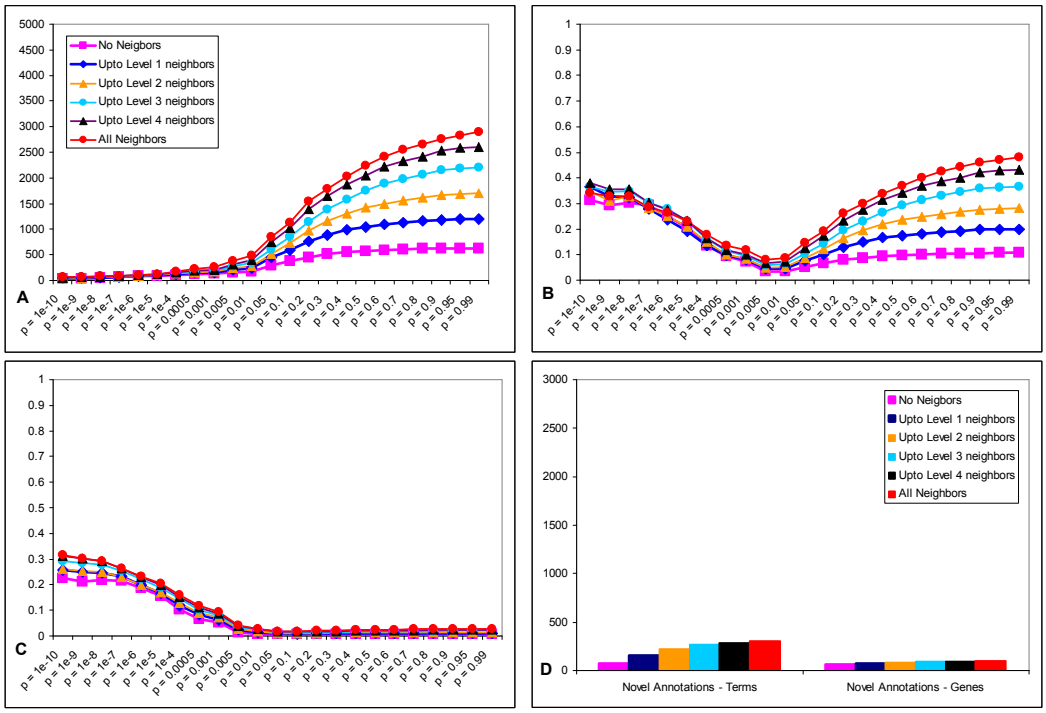

Figure 7.5. Expression correlation (ρ > 0.90), molecular function terms

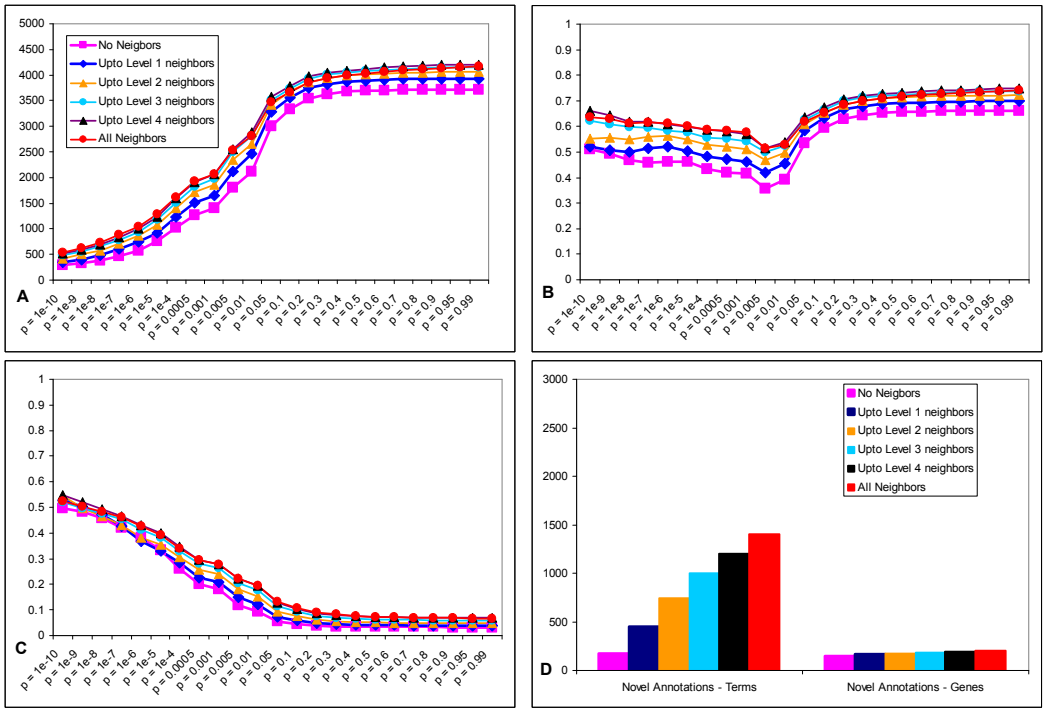

Figure 7.6. Homology, biological process terms

The results obtained are given in Figures 7.6 and 7.7. Number of predicted ORFs increase as the threshold of p-value and number of included neighbors are increased as in correlation network cases. The fraction of correctly predicted terms decrease as the threshold of p-value is increased. However the fraction of predicted ORFs with at least one correct prediction remains above 0.4 with almost all cases proving high prediction power. Moreover, the number of ORFs that can be predicted is high, for selected parameters, (4 neighbors, $p < 1 \times 10^{-7}$), the number of ORFs with at least one correctly predicted terms is 831 and 1117 for biological process and molecular function terms respectively. For 190 and 253 unknown ORFs, 1206 and 938 biological process and molecular function terms are predicted respectively. Fractions of ORFs with at least one correctly predicted are expected to be 0.61 and 0.70, while fractions of correct predictions are expected to be 0.46 and 0.50 correspondingly.

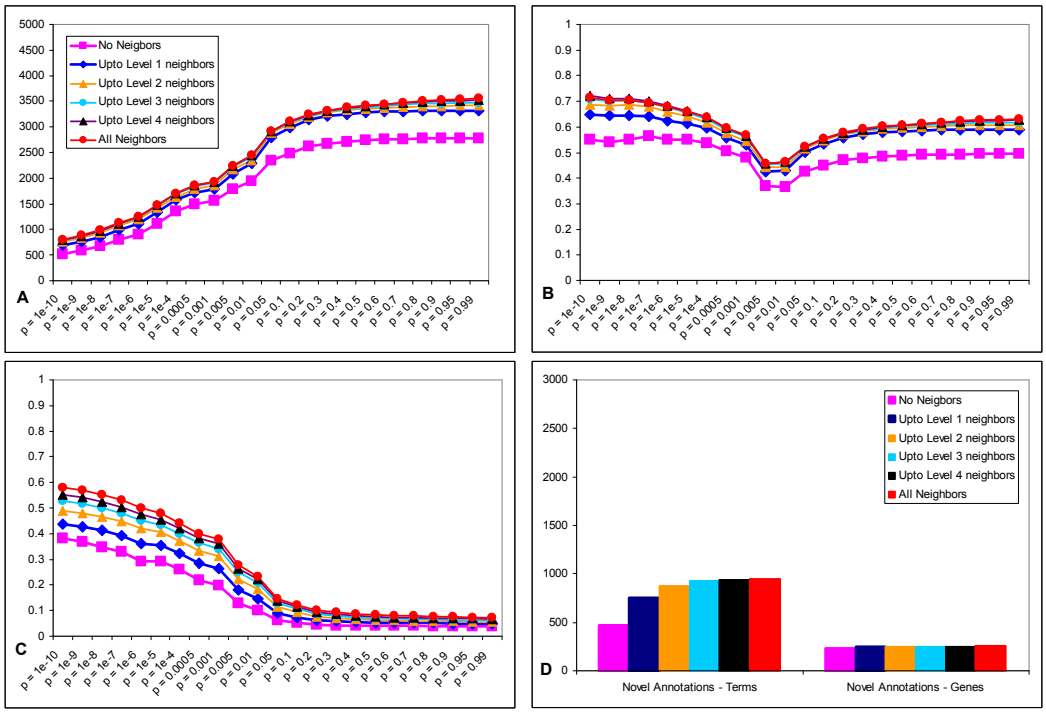

Figure 7.7. Homology, molecular function terms

7.2.3. Prediction Power of Interactome Data on GO Classification Trees

The interaction network was constructed using the data from BIND, STRING and GRID databases. The power prediction analysis was carried out using 23 levels of

threshold p-values and 6 levels of neighborhood to infer the prediction power of interaction network on GO prediction tree of biological process terms.

The results obtained are given in Figures 7.8 and 7.9. Similar to previous analyses, increasing the threshold p-value and the number of neighbors included, increases the number of ORFs predicted. The fractions of the correctly predicted ORFs and terms follow a similar trend, and the values are higher, indicating that interaction data have more prediction power on GO prediction trees when compared to other biological networks. Expected fractions of ORFs with at least one correctly predicted term 0.82 and 0.67 and expected fractions of correctly predicted terms 0.68 and 0.65, these values are highest fractions obtained among all cases using 4 level of neighbors and $p < 1 \times 10^{-7}$. Protein-protein interaction network allows prediction of 1373 and 843 ORFs with at least one correctly predicted terms (4 neighbors, $p < 1 \times 10^{-7}$). 1732 and 1126 biological process and molecular function terms predicted for 1126 and 242 unknown ORFs respectively.

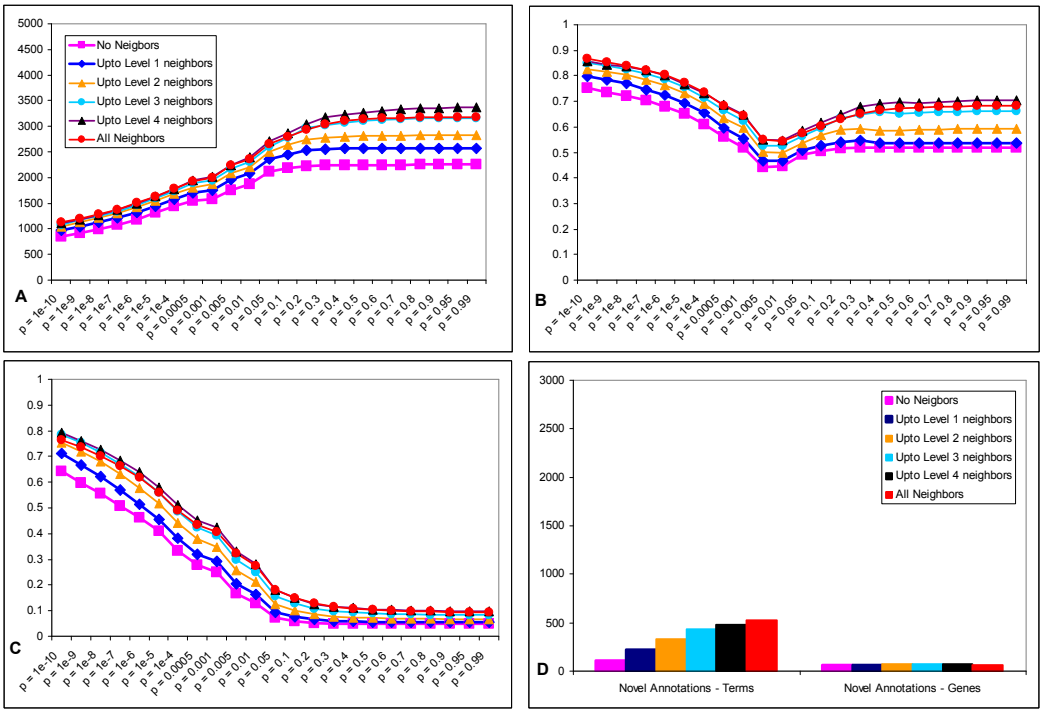

Figure 7.8. Interactome, biological process terms

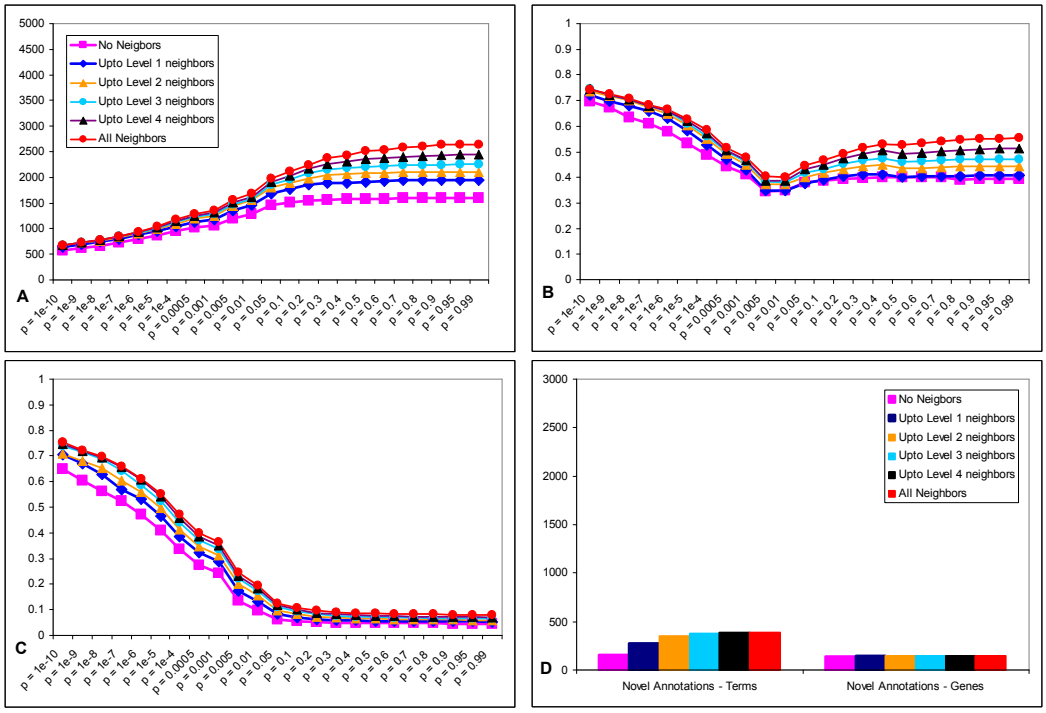

Figure 7.9. Interactome, molecular function terms

7.2.5. Prediction Power of Regulatory Network on GO Classification Trees

The gene pairs bound by at least same two TFs were connected to construct the regulatory network. The prediction power of this network on GO biological process and molecular function term trees were tested using using 23 levels of threshold p-values and 6 levels of term neighborhood (Figures 7.10 and 7.11).

Trends of the number of ORFs predicted, fractions of correctly predicted ORFs and terms are similar to previous cases. However, this network allows prediction of a lower number of ORFs since the number of ORFs included in the network is small (2253) when compared to other networks. Also, the number of ORFs that can be predicted is low, for selected parameters, (4 neighbors, $p < 1 \times 10^{-7}$), the number of ORFs with at least one correctly predicted terms is 656 and 458 for biological process and molecular function cases respectively. 70 and 147 unknown ORFs are predicted to be predicted by 483 and 384 biological process and molecular function terms respectively.

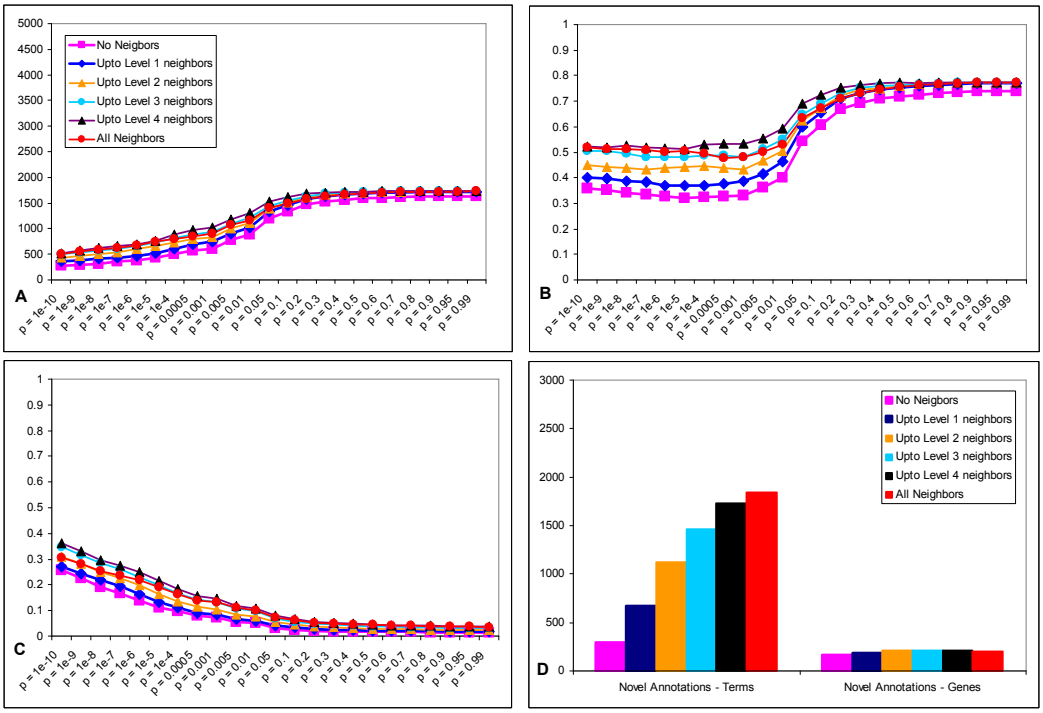

Figure 7.10. Regulatory network, biological process terms

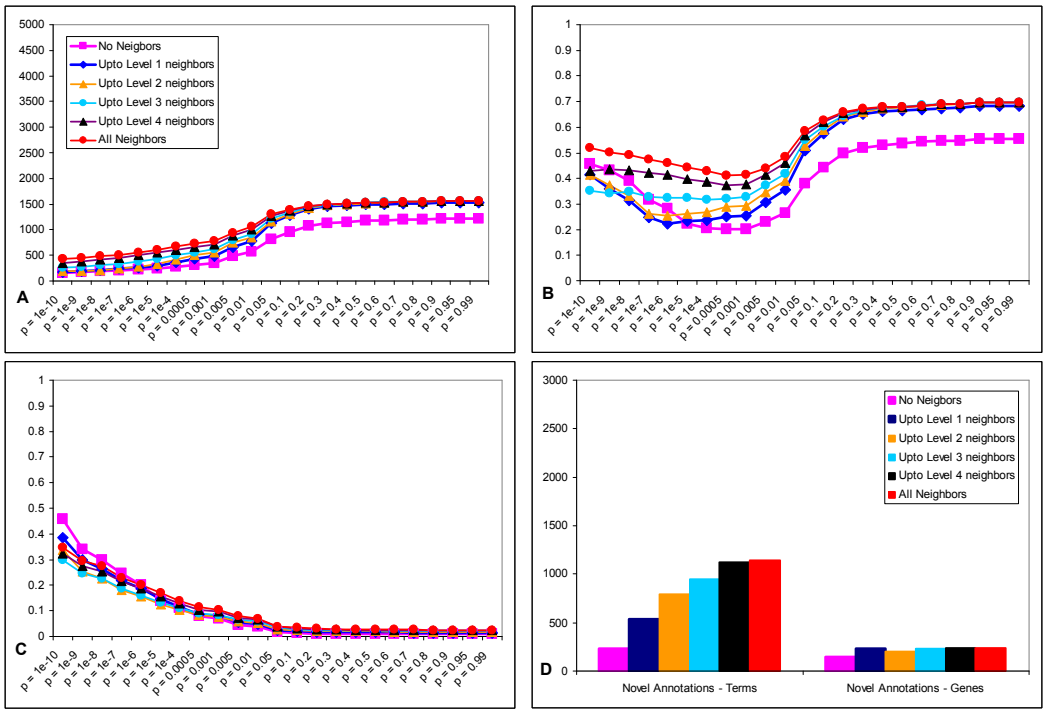

Figure 7.11. Regulatory network, molecular function terms

7.2.5. Prediction Power of Compendium of Networks on GO Classification Trees

The edges and nodes of all biological networks constructed were combined to obtain a larger network which concerns all the biological information available. The multiple edges between a pair of ORFs were reduced to one, thus every pair of ORFs is connected by only one edge at most. In total, 6702 ORFs were included in this network with a total of 631510 edges.

The analyses were repeated for the compendium of networks using both biological process and molecular function terms. The results are given in Figures 7.12 and 7.13. Similar trends were observed when compared to previous cases, that is, the number of genes correctly predicted increase, the fraction of correctly annotated terms decrease and number of novel annotations increase as the p-value threshold is increased. The fraction of correctly predicted ORFs with at least one term decreases as the p-value threshold approaches to 0.01 and increases for higher or lower thresholds.

The compendium of networks allows more ORFs to be correctly predicted with at least one term. The fractions of correct predictions are generally high when compared to single biological networks. For selected parameters, (4 neighbors, $p < 1 \times 10^{-7}$), 2213 and 1747 ORFs with at least one correctly predicted terms were found for biological process and molecular function terms respectively. 443 and 524 unknown ORFs are predicted to have 2981 and 1975 biological process and molecular function terms respectively. Expected fractions of ORFs with at least one correctly predicted term are 0.68 and 0.63 for biological process and molecular function cases while the expected fraction of correctly predicted terms were 0.45 for both cases.

Both the number of predictions and expected fractions of correctness are higher in the compendium of networks when compared to the results of single networks indicating that combining all biological information increases the prediction power. The expected fraction of correctness is higher in compendium of networks when compared to results of single networks, except for protein amino-acid sequence similarity and protein-protein interaction networks. The number of predictions is higher in the compendium of networks when compared to all other networks investigated. These findings indicate that combining

all biological information generally increases the prediction power and allows more predictions to be made.

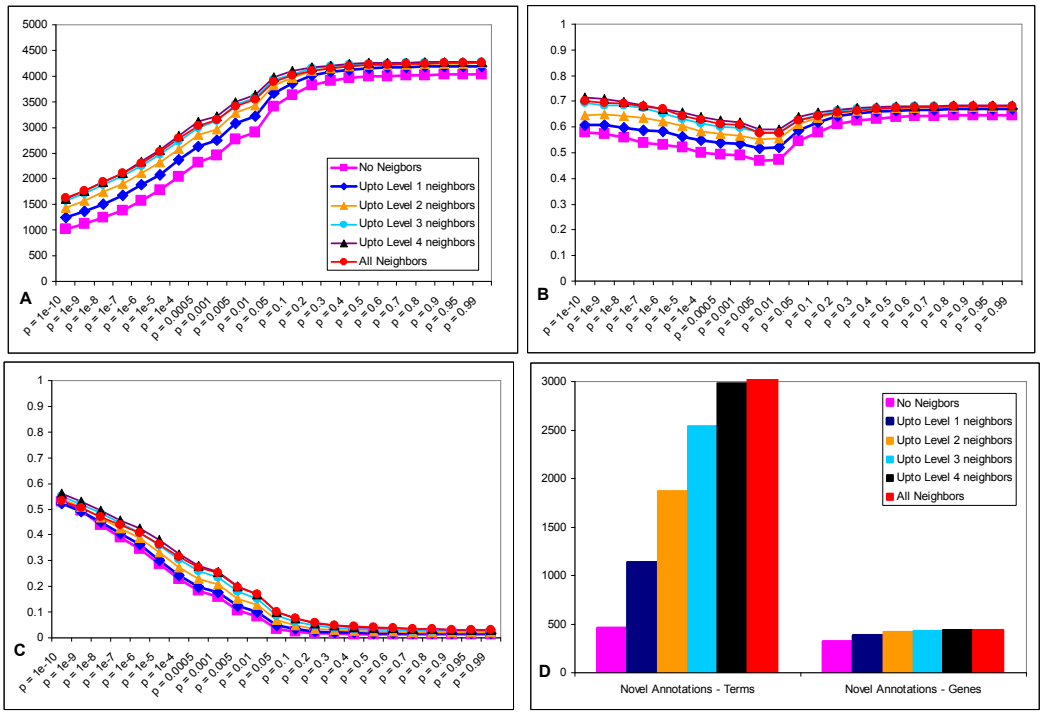

Figure 7.12. Compendium network, biological process terms

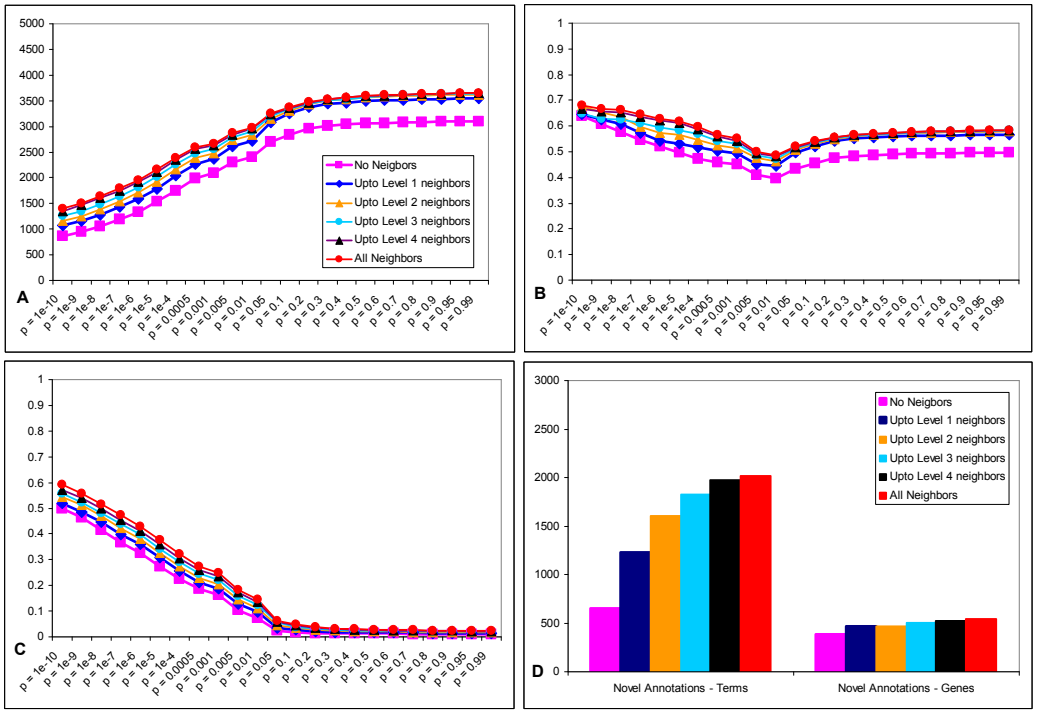

Figure 7.13. Compendium network, molecular function terms

7.2.6. Prediction Power of All Networks for the Selected Parameters

The threshold p = 1 x 10^{-7} was selected as an optimum threshold since it allows satisfactory number of ORFs to be predicted with acceptable fractions of correctness. The level of neighbors (or parents) to be included analyses were selected as 4 since at this level adequate number of parents from the classification tree are included while the most general terms are excluded. In Figures 7.14 and 7.15 a comparison of all networks are given where the predictions were made using 4 levels of neighbors. A summary of statistics on predictions made using p < 1 x 10^{-7} and 4 level of neighbors are given in Table 7.4.

The expression correlation networks and regulatory network generally allow a smaller number of ORFs to be predicted than the other networks. Moreover, the fraction of correct predictions are lower for these networks at significant p-value thresholds. Thus the prediction power of these biological networks is lower when compared to other networks. Homology and interactome networks enable more ORFs to be predicted with higher fractions of correctness. Their fraction of correctness remains high at almost all p-value thresholds. It can be concluded that these networks are more informative than the former ones.

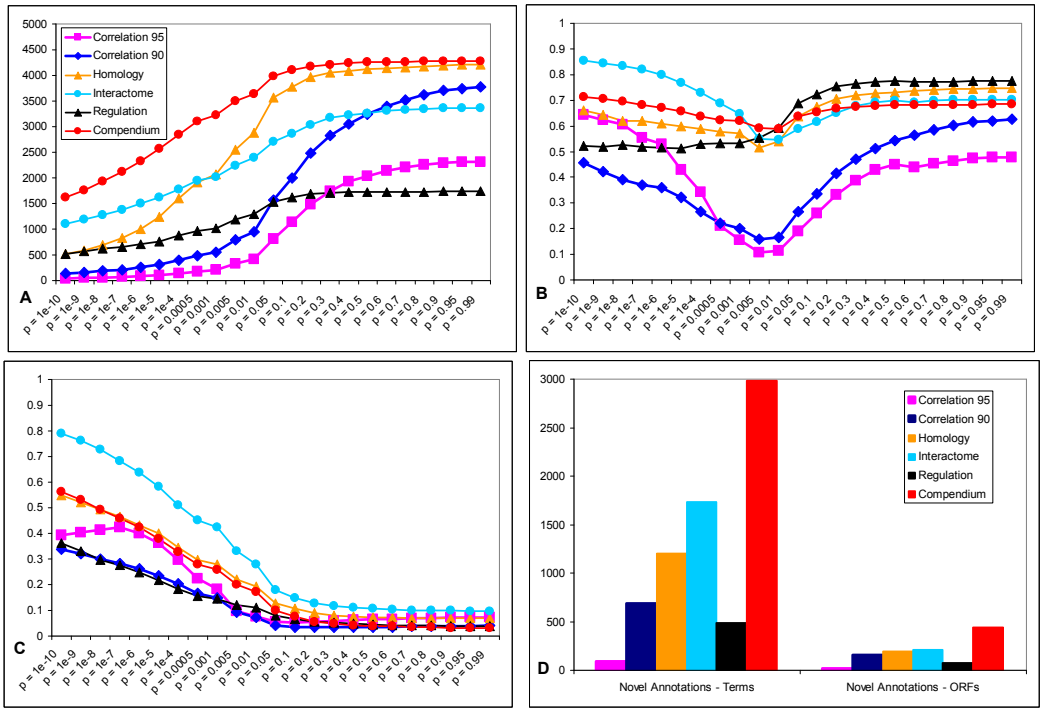

Figure 7.14. Comparison of all networks including 4 neighbors, biological process terms

Table 7.4. Statistics for $p < 1 \times 10^{-7}$ and 4 level of neighbors

Biological Process	Expression Correlation ($\rho > 0.95$)	Expression Correlation ($\rho > 0.90$)	Homology	Interactome	Regulaiton	Compendium
Novel Annotations - Terms	89	686	1206	1732	483	2981
Novel Annotations - ORFs	17	154	190	210	70	443
Fraction of ORF with at least one correctly predicted term/ORFs predicted	0.55	0.37	0.61	0.82	0.52	0.68
Fraction of correctly predicted terms/predicted terms	0.42	0.28	0.46	0.68	0.27	0.45
Number of ORFs with at least one correctly predicted term	69	214	831	1373	656	2113
Molecular Function	Expression Correlation ($\rho > 0.95$)	Expression Correlation ($\rho > 0.90$)	Homology	Interactome	Regulaiton	Compendium
Novel Annotations - Terms	44	289	938	1126	384	1975
Novel Annotations - ORFs	11	93	253	242	147	524
Fraction of ORF with at least one correctly predicted term/ORFs predicted	0.54	0.30	0.70	0.67	0.42	0.63
Fraction of correctly predicted terms/predicted terms	0.42	0.26	0.50	0.65	0.21	0.45
Number of ORFs with at least one correctly predicted term	34	85	1117	843	458	1747

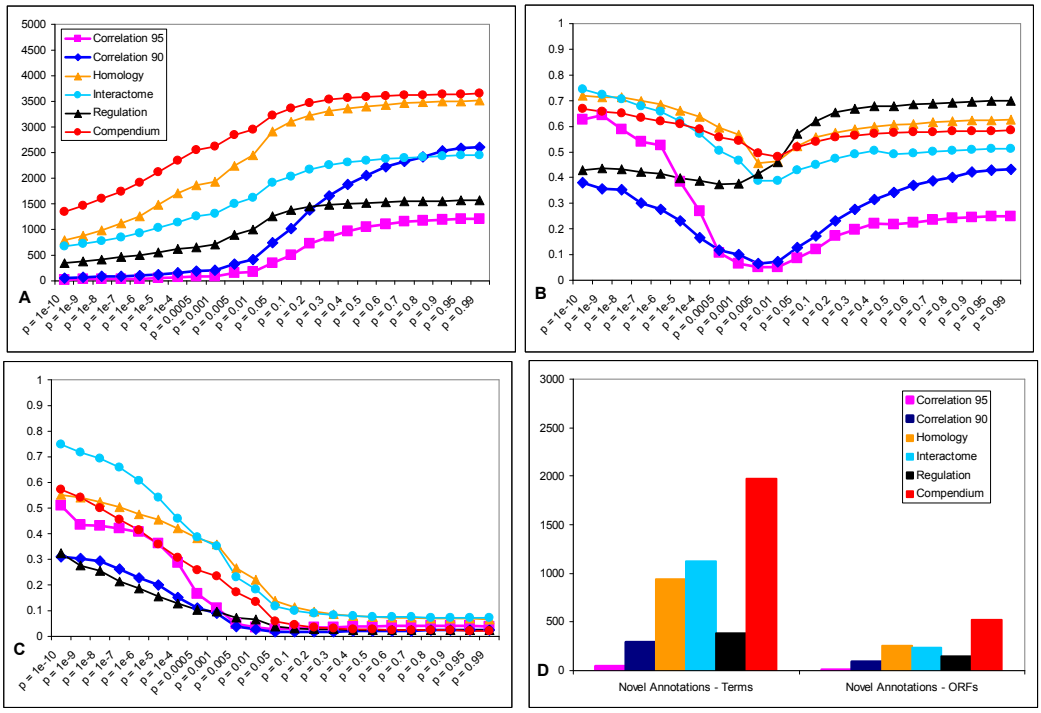

Figure 7.15. Comparison of all networks including 4 neighbors, molecular function terms

The best prediction power is found in compendium of networks. First of all, number of ORFs correctly predicted by at least one term is largest for the compendium in both biological process and molecular function predictions. Fraction of correct predictions in compendium network is generally higher than the other networks for most cases. This network also allows maximum number of unknowns to be predicted with high fraction of correctness. The ORFs and biological process and molecular function terms predicted using parameters as 4 neighbors, $p < 1 \times 10^{-7}$ are given in Table E.1 and E.2, respectively.

8. CONCLUSIONS AND RECOMMENDATIONS

In this section, the concluding remarks on the findings of this study are presented and future directions for answering the questions and testing the hypotheses that arise as a result of this study are given.

8.1. Conclusions

All aspects of metabolome and transcriptome data obtained in chemostats were investigated in the present study. The conclusions drawn from the analyses in Sections 4-7 are presented below in separate subsections.

8.1.1. Transcriptional Regulation of Central Carbon Metabolism

Yeast *S. cerevisiae* has a flexible central carbon metabolism which is capable of adapting to a wide range of carbon sources and environmental conditions. All perturbations applied in this study target the central carbon metabolism of yeast, thus the detailed analysis of gene expression was focused on central pathways. The results verify the transcriptional regulation information provided in the literature and enable new regulatory relations to be proposed.

Homozygous deletion mutants *hap4Δ/hap4Δ*, *bcs1Δ/bcs1Δ*, *oxa1Δ/oxa1Δ* and *rip1Δ/rip1Δ* are completely respiratory deficient. Heterozygous deletion mutants *HAP4*/*hap4Δ*, *RIP1*/*rip1Δ* and homozygous deletion strain *mba1Δ/mba1Δ* have partial respiratory deficiency. Homozygous and heterozygous deletion strains m*ig1Δ/mig1Δ* and *MIG1/mig1Δ* have no growth deficiency under the conditions investigated.

Hap4p and Mig1p are known transcriptional regulators. *RIP1* and *MBA1* have Hap4p binding sites in their upstream regions. Results of this study indicated that, in addition to *RIP1* and *MBA1*, *MIG1* may also be regulated by Hap4p; however, it has no Hap4p binding site. Both *HAP4* and *MBA1* were observed to be transcriptionally affected in

mig1/mig1Δ and *MIG1/mig1Δ* deletion mutants. *HAP4* has Mig1p binding site but *MBA1* does not.

Mba1p is not generally recognized as a transcriptional factor; however, in this study, many genes from central carbon metabolism were upregulated in *mba1/mba1Δ* mutant, in addition to being upregulated in *mig1Δ/mig1Δ* and *MIG1/mig1Δ* deletion mutants. The *MIG1* gene was also downregulated in *mba1Δ/mba1Δ* mutant, thus the results obtained indicate that Mba1p may also be a transcriptional regulator.

RIP1, *OXA1* and *BCS1* are also not recognized as transcription factors; however, in this study, many genes were observed to be upregulated in deletion mutants of these genes. Deletions of these genes cause respiratory deficiency, which results in activation of retrograde signalling; thus, the genes affected in these deletion mutants may be upregulated as response to retrograde regulation. Mig1p and Mba1p may be repressing both ethanol production and respiration because their deletion results in upregulation of most of the central carbon metabolism pathways.

The results of this study show that, in a *hap4Δ/hap4Δ* deletion mutant, the nuclear genes encoding the components of respiratory chain are downregulated. These results confirm that Hap4p is an activator of respiration-related genes. The mitochondrially encoded respiratory chain genes are downregulated by *oxa1Δ/oxa1Δ* deletion mutant. The present results indicate that Oxa1p may be needed for transcription of mitochondrially encoded respiratory chain components.

Respiration deficiency also upregulates expression of genes from most of the central pathways, this may be an attempt by the cells to adapt to this severe deficiency.

Ammonium limitation has a widespread effect on central metabolism, probably because of the high glucose concentration in ammonium-limited medium. The respiratory chain is generally downregulated while galactose metabolism and glycolysis are upregulated. Production of acetaldehyde from pyruvate is also upregulated which may indicate that glucose repression causes the carbon source to be utilized by fermentation.

Upregulation of various genes encoding hexose transporters, proteins acting on the glucose-signaling, glycolysis, and galactose pathways by *hap4Δ/hap4Δ, HAP4/hap4Δ* and *RIP1/rip1Δ* deletions were observed. These genes are not affected in homozygous deletion mutant *rip1Δ/rip1Δ*. The biological reason of this expression behaviour remains unclear.

The detailed investigation of transcriptional regulation of central carbon metabolism has resulted in interesting findings. Experiments employing molecular biology techniques need to be conducted to support the present results and to explain the reasons behind the expression behaviour observed.

8.1.2. Integration of Transcriptome and Metabolome Data

The systems biology approach demands that data from high-throughput technologies to be analysed together to obtain a complete picture of cell behaviour. In the present study, bioinformatics tools were used to reduce the dimensionality of high throughput transcriptome and metabolome data towards pattern recognition and data integration.

Linear modelling enables filtering of the insignificant variation in data sets, and PCA allows clustering of the samples with similar mRNA expression or metabolite profiles.

The PLS method was applied to metabolic and transcriptomic data to gain insight into the changes in metabolism due to three factors (growth medium, dilution rate and gene deletion). The method enabled extraction of the following information from these data sets:

1. Discrimination of the effects of the above factors on transcriptome and metabolic data.
2. Modelling of metabolic data as a function of transcriptome data and elucidation of the extent of correlation between these two data sets.
3. Identification of ORFs that mediate the changes in metabolic data in response to perturbations.

Investigation of variation in metabolic and transcriptomic data in response to perturbations in growth media, growth rate and gene deletions resulted in elucidation of

pathways that are mediating the effects of the perturbations on the metabolic data. The genes from biological processes “cellular respiration”, “phosphorylation”, “coenzyme biosynthesis” and “nucleotide metabolism” are significant in modelling of metabolic data, thus, activation of these processes are thought to be important sources of variation in the metabolic profiles of the cells.

Investigation of transcriptome data in combination with metabolic data and extracellular metabolome profiles yielded significant findings. The metabolic profile, i.e., biomass content, glucose consumption and ethanol production rates of cells, depends on the respiration sufficiency of the cells. In respiratory-sufficient cells, “ion transport” and “metal ion homeostasis” processes were active while “organic acid metabolism” and “fermentation” were active in respiratory-deficient cells.

The complete extracellular metabolome profiles enable more detailed investigation of the effects of the gene deletions. The mutant *hap4Δ/hap4Δ* was an outlier in PLS analysis of transcriptome and extracellular metabolome data. This is an expected behaviour since the expression profile of central carbon metabolism genes were different in mutant *hap4Δ/hap4Δ* when compared to all other deletion mutants (Section 4). However, this behaviour cannot be identified when PLS is applied to transcriptome and metabolic data.

In Section 4, it was explained that respiratory-deficient mutants other than *hap4Δ/hap4Δ* have shared effects on gene expression and similarly the mutants *mba1Δ/mba1Δ*, *mig1Δ/mig1Δ* and *MIG1/mig1Δ* have shared effects on gene expression. In mutants other than *hap4Δ/hap4Δ*, "oxidative phophorylation” and “nucleotide biosynthesis” is upregulated. This result indicates that *hap4Δ/hap4Δ* deletion downregulates the respiration and (probably) growth-related activities. Regulation of metabolism was upregulated in mutants *rip1Δ/rip1Δ*, *RIP1/rip1Δ*, *HAP4hap4Δ*, *bcs1Δ/bcs1Δ*, and possibly in *oxa1Δ/oxa1Δ* mutants, while “cell redox homeostasis” and “response to oxidative stress” were upregulated in *mig1Δ/mig1Δ*, *MIG1/mig1Δ* and *mba1Δ/mba1Δ* mutants. Activation of these processes indicate that the cells bearing these gene deletions are forced to activate the pathways for adapting to gene deletion and re-balancing the metabolism.

Even in cases where the number of variables in the metabolic/metabolome data is much higher, the PLS method would enable the identification of metabolites that are affected by the conditions applied as well as the genes that mediate the effects of the conditions. In the present dataset, the metabolome data were available as a "signature" where the individual metabolites cannot be identified. However, valuable information on adaptation of metabolism to perturbations was deduced even without identification of individual metabolites.

Finally, unknown genes can be annotated using this methodology and studies towards product maximization can be conducted by identifying the genes and the pathways that are responsible for the changes in formation of metabolic products.

8.1.3. Annotation of Unknown Yeast ORFs

In this study, changes in the expression of genes were used to elucidate the metabolic pathways and regulatory mechanisms that respond to environmental or genetic modifications. Results from previously published chemostat data sets were merged with novel data generated in the present study.

Transcriptional co-regulation of genes may indicate shared functions or "guilt-by-association". Expression of ORFs displaying significant changes in correlation with other ORFs were analysed using GO mapping tools and supplemented by literature information.

Annotations for 17 of the investigated ORFs were proposed and 7 of these annotations were supported by external evidence. For the 10 ORFs, the annotations made depend solely on transcriptome data supported by weak, or no, external evidence. ORFs *YJL103c*, *YLR327c, YGR243w, YGR067c*, *YMR090w*, *YOR173W (DCS2)*, *YKL187c*, *YGR236c (SPG1)*, *YER067w*, *YGR173W*, *YFR017c*, *YER121w*, *YDR070C (FMP16)*, *YGL157w*, *YMR107w (SPG4)*, *YMR206w*, *YNL175C (NOP13),* and *YJL200C* were assigned functions as a result of this study.

YJL103c was assigned the process "transcriptional regulation of respiration" ($p < 10^{-13}$), with external evidence. *YLR327c* was assigned the process "glycogen and energy

reserve metabolism ($p < 10^{-10}$) and "transferase" function ($p < 10^{-4}$) with external evidence. *YGR243w* was assigned the process "energy derivation and carbohydrate metabolism" ($p < 10^{-6}$) with external evidence. *YGR067c* was assigned the process "generation of precursor metabolites and energy" ($p < 10^{-6}$) with weak external evidence. *YMR090w* was assigned the process "glutamate synthesis pathway" ($p < 10^{-5}$) with external evidence. *YOR173W (DCS2)* was assigned the process "carbohydrate and trehalose metabolism" ($p < 10^{-4}$) with weak external evidence. *YKL187c* was assigned the process "carbohydrate and alcohol metabolism" ($p < 10^{-4}$) without any external evidence. *YGR236c (SPG1)* was assigned the process "water deprivation" ($p < 10^{-4}$) without any external evidence. *YER067w* was assigned the process "energy derivation metabolism" ($p < 10^{-4}$) without any external evidence. *YGR173W* was assigned the process "purine salvage, tRNA modification and ribonucleoside monophosphate biosynthesis pathways" ($p < 10^{-4}$) without any external evidence. *YFR017c* was assigned the process "generation of precursor metabolites and energy" ($p < 10^{-3}$) with external evidence.

YER121w was assigned the processes "generation of precursor metabolites and energy" ($p < 10^{-6}$) and "phosphatase regulation" ($p < 10^{-3}$) with weak external evidence. *YDR070C (FMP16)* was assigned the process "response to water" ($p < 10^{-3}$) with weak external evidence. *YGL157w* was assigned the process "transport and carboxylic ester hydrolase" ($p < 10^{-3}$) without any external evidence. *YMR107w (SPG4)* was assigned the function "intra-golgi transport and v-SNARE activity" ($p < 10^{-3}$) without any external evidence. *YMR206w* was assigned the process "regulation of carbohydrate biosynthesis" ($p < 10^{-3}$) without any external evidence. *YNL175C (NOP13)* and "ribosome biogenesis and RNA metabolism" ($p < 10^{-3}$) with external evidence. *YJL200C* was assigned the process "transcription from RNA polymerase I promoter" ($p < 10^{-3}$) without any external evidence.

It should be kept in mind that the predictions made with low-precision p-values should be considered only if they are supported by external evidence. All annotations with $p < 10^{-5}$ were supported by such evidence, and so this value may be considered as a safe cut-off limit for such annotations.

8.1.4. Investigation of Prediction Power of Networks for Functional Annotation

In silico methods for functional annotation of unknown genes frequently make use of the assumption "similaries among the properties of the genes may indicate shared functions". In identification of functions of unknown genes, similarities in genomewide data on gene expression, gene product localization, structural/sequence homology, transcription factor binding, phenotypes of deletion mutants and physical/genetic interactions among proteins/genes proved to be useful.

In this study, a method is proposed to assess the prediction power of biological networks in assigning classes to unknown ORFs using the information on genes with known classes. The developed method can be used for any network composed of genes presumed that some portion of the genes in the network are already annotated to some classes. The method was used to investigate the networks of ORFs with correlated expression, protein homology, transcription factor binding data and physical/genetic interactions. SGD GO terms were used as the classes. 2981 biological processes and 1975 molecular functions for 443 and 524 unknown ORFs, respectively, are predicted as an outcome of combining all biological information. Expected fractions of ORFs with at least one correctly predicted term are 0.68 and 0.63 for networks of biological process and molecular function terms, respectively; expected fraction of correct predictions is 0.45 for both of the classification networks. The software PredPower is developed to carry out all computations in Matlab environment.

8.2. Recommendations

This study was focused on gaining new information on respirative and fermentative central carbon metabolism of yeast *S. cerevisiae* and functional annotation of unknown ORFs. The results propose novel transcriptional regulation relations among many genes and novel functions to more than 500 unknown genes.

These results are elucidated using data from high throughput technologies, which provide only a general picture of the cell state. Thus, these results produces hypotheses and

questions that need to be tested and answered by more standart molecular biology techniques:

1. Does Oxa1p regulate transcription of mitochondrially encoded components of respiration chain?
2. How is the coordinated effect of Mig1p and Mba1p on central carbon metabolism mediated?
3. How is the effect of respiration deficiency conducted to the whole central metabolism?
4. Why are the genes encoding hexose transporters affected from heterozygous deletion of the *RIP1* gene, but not from its homozygous deletion?

Proteome is another high throughput data type, which provides the profile of protein content in cells. A verification method for transcriptional regulation could be collecting proteome data under the conditions investigated. Also an emerging technology, protein arrays, could be used to detect the proteins with specific properties to verify the present results.

17 genes were annotated novel functions in this study. The annotated functions need to be verified using various methods, such as conducting fermentations using deletion mutants of these genes to observe their phenotypes.

The PredPower tool, which was generated as another output of this study, can be improved in many ways. First of all, a user interface is needed for convenience of users. The networks analysed by the tool may be investigated after assigning weights to the edges or nodes. More options for evaluating the results may be included.

This section concludes the results obtained in this study. It is strongly believed that this study provides valuable novel biological information and new methodologies to the systems biology research community that aims to solve the yeast “puzzle”.

APPENDIX A: NOTES ON EXPERIMENTS

In this section, the additional notes on the experimental part of this study will be provided and results of the supplementary experimental work will be given.

A.1. Location of the Experiments

The majority of the experiments were performed in Faculty of Life Sciences, The Michael Smith Building, University of Manchester, Oxford Road, Manchester, M13 9PT United Kingdom between dates June-October 2003 and June-October 2004. Additional analyses were conducted in Department Chemical Engineering Department, KB440, Boğaziçi University, Bebek, 34342 Istanbul Turkey.

A.2. Verification of Gene Deletions

The strains were tested for correct deletion using PCR-based methods. Locations of the primers on the genome are given in Figure 3.1 and sequences of primers are given in Table 3.2. *Kan*MX forward and reverse primers were used to verify the existance of *Kan*MX casette on the genome. Uptag-Downtag primers are used to verify the existance of correct deletion casette and A-D primers are used to verify that deletion casette is in correct location on the genome. Upon completion of fermentations, DNA was isolated from chemostat samples to verify the growth of correct mutants in each reactor.

Only the PCR products of A-D primers and DNA samples from the chemostats will be given here, which verifies that DNA segment with expected length exist in the locus of gene under interest. The expected length of PCR products for wildtype alleles and alleles carrying deletions are given in Table A.1. For homozygous deletion mutants, single band with “deletion” length is expected, for homozygous deletion mutants, two bands are expected with length “deletion” and “wildtype”.

In Figure A.1, the ladder BioLine HyperLadder IV used to estimate the length of PCR products of A-D primers is given. All chemostat samples were successfully verified

to be the correct strain, however, no band is observed for *MIG1* deletion strains (Fig.A2). It is highly probable that the A-D primers for *MIG1* were not working. Existance of correct deletion in these strains were verified using the array data, together with all the other deletions (results not shown).

Table A.1. Expected lengths of PCR products of AD primers (bp).

Gene	Deletion	Wildtype
HO	2230	2407
RIP1	2280	1344
HAP4	2239	2320
BCS1	2282	2069
MIG1	2069	2000
MBA1	2251	1504
OXA1	2221	1846

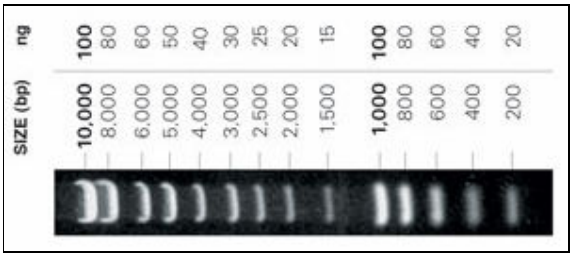

Figure A.1. Ladder used for estimation of length of PCR products

A.3. Respiration Tests

Only respiration-competent cells can grow on non-fermentable carbon source ethanol. The strains used in this study were tested for respiratory deficiency on YPE plates (2 per cent yeast extract, 1 per cent peptone, 2 per cent ethanol). One or two colonies were taken from a YPD plate, suspended in 140 microliter of sterile dH_2O, and the suspension was diluted with sterile dH_2O (1:8) five times. 2 microliter of 6 suspensions at different concentrations were spotted on YPE plates. Figure A.3 illustrates the growth of strains after incubation at 30^oC for three days.

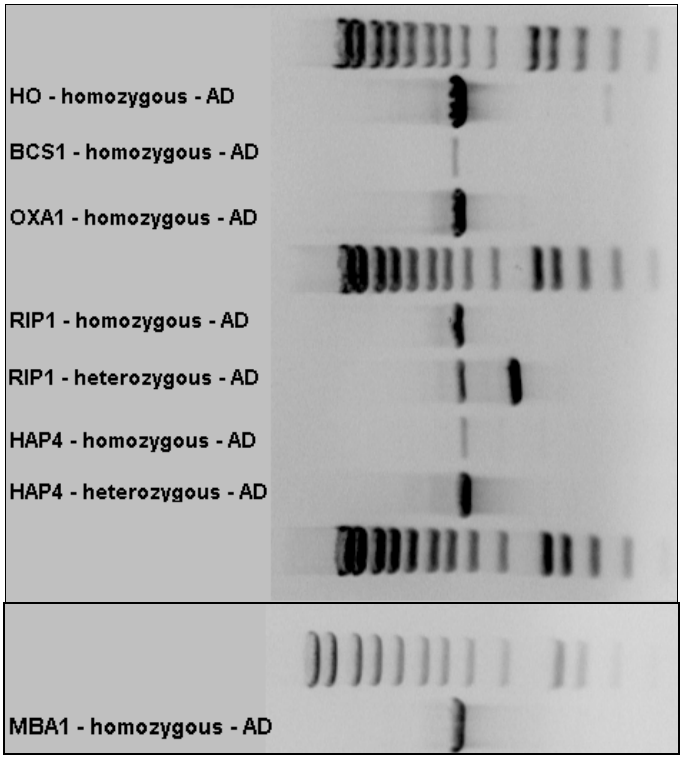

Figure A.2. PCR products of A-D primers and DNA from chemostat samples.

Almost all heterozygous strains grow as fast as the reference strain *hoΔ/hoΔ*. Respiratory deficiency is apparent for homozygous deletion strains *hap4Δ/hap4Δ*, *bcs1Δ/bcs1Δ*, *mba1Δ/mba1Δ* (partial deficiency), and *oxa1Δ/oxa1Δ* and heterozygous deletion mutant *RIP1*/*rip1Δ* (partial deficiency). These results are in agreement with results of Steimetz *et al.* (2002) except for homozygous *rip1Δ/rip1Δ* and heterozygous deletion *HAP4*/*hap4Δ* mutants. Homozygous deletion strain *rip1Δ/rip1Δ* was expected to be respiratory deficient, but it grows on YPA-E plates. The test was repeated for fresh homozygous *rip1Δ/rip1Δ* and heterozygous deletion *HAP4*/*hap4Δ* mutant cells from the library and complete respiratory deficiency for *rip1Δ/rip1Δ* and partial respiratory deficiency for *HAP4/hap4Δ* were observed (Figure A.4). The deletion tests were also repeated for these strains and deletion-verified cells were used in chemostat experiments.

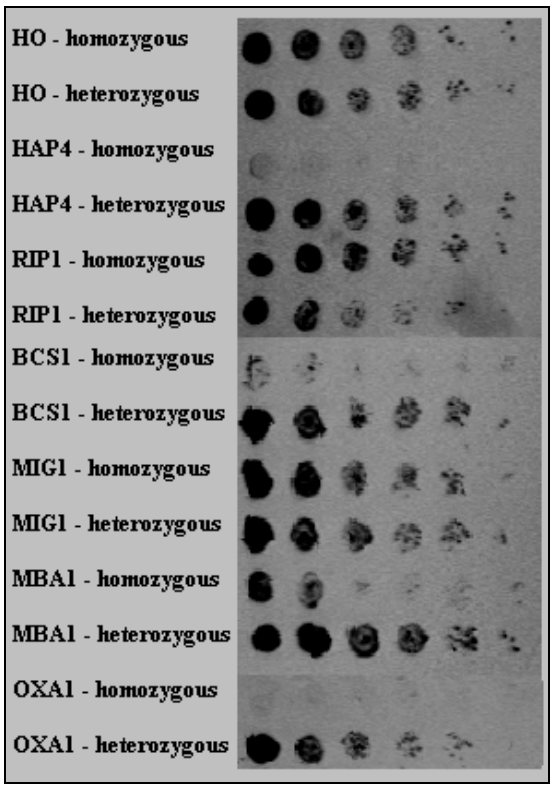

Figure A.3. Respiration test on YPA-E plates

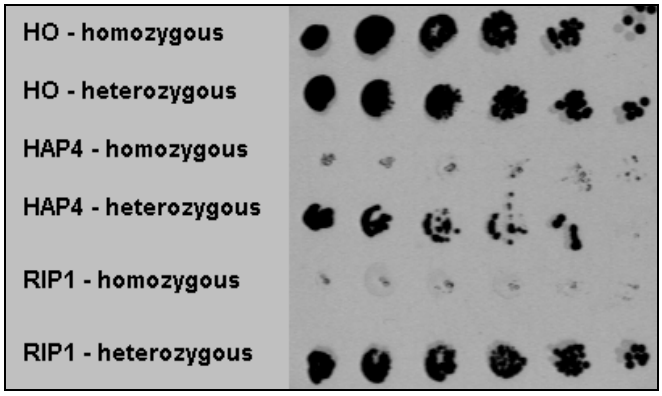

Figure A.4. Respiration test repeated on YPA-E plates

APPENDIX B: SUPPLEMENTARY TABLES FOR SECTION 6

In this section, the tables mentioned in Section 6 (Annotation of Unknown Yeast ORFs) are provided. Table B.1 can not be given here explicitly since its dimensions are not appropriate for printing, only few representative rows are given. Complete Table B.1 is given in CD version of this thesis, saved under file name TableB1.xls.

Table B.1. 638 significantly affected ORFs, effects of factors and p-values

ORF	Gene	Deletion effect	Medium effect	Dilution Rate Effect	P (Deletion)	P (Medium)	P (Dilution Rate)
YFR015C	GSY1	2.49E-01	-2.32E+00	-2.38E-01	1.36E-01	6.49E-05	1.49E-01
YHR094C	HXT1	-1.98E-01	4.55E+00	-3.89E-02	5.94E-01	1.85E-04	9.15E-01
YNR059W	MNT4	2.46E-02	5.13E-01	-2.40E-01	5.68E-01	2.03E-04	3.71E-03
YDR345C	HXT3	9.68E-02	3.38E+00	-7.98E-02	7.37E-01	2.29E-04	7.81E-01
YKL093W	MBR1	-4.16E-02	-1.88E+00	-6.63E-01	7.96E-01	2.36E-04	1.15E-02
YMR081C	ISF1	2.48E-01	-3.02E+00	-9.95E-01	3.80E-01	2.79E-04	1.68E-02
YPR184W	GDB1	9.33E-02	-2.05E+00	-4.67E-01	6.45E-01	3.98E-04	6.72E-02
YBR297W	MAL33	3.20E-01	-2.60E+00	-2.86E-01	2.67E-01	4.74E-04	3.14E-01
YER028C	YER028C	-1.56E-01	2.76E+00	9.78E-03	5.90E-01	4.84E-04	9.72E-01
YKR058W	GLG1	1.05E-01	-7.56E-01	-3.04E-01	2.88E-01	9.10E-04	2.37E-02
YJL166W	QCR8	-8.11E-01	-6.85E-01	-1.85E-01	5.31E-04	1.02E-03	8.17E-02
YLR273C	PIG1	7.72E-02	-6.32E-01	-3.08E-01	3.56E-01	1.03E-03	1.41E-02
YDL157C	YDL157C	-2.64E-01	-6.31E-01	8.72E-02	2.48E-02	1.11E-03	3.11E-01

Table B.2. Conditions covered by experiments

Condition	Experimental Conditions	Covered by Experiments
UN	Upregulation at nitrogen/ammonium limitation/starvation*	3, 5 and 6
DN	Downregulation at nitrogen/ammonium limitation/starvation	3, 5 and 6
UC	Upregulation at carbon/glucose limitation/starvation	3, 5 and 6
DC	Downregulation at carbon/glucose limitation/starvation	3, 5 and 6
DH	Downregulation at HAP4 deletion mutants	1 and 4
UR2-DR1	Upregulation at higher dilution rates	1 and 7
DR2-UR1	Downregulation at higher dilution rates	1 and 7
UN-DC *	Upregulation at nitrogen/ammonium limitation with respect to carbon/glucose limitation	1 and 2
DN-UC ‘	Downregulation at nitrogen/ammonium limitation with respect to carbon/glucose limitation	1 and 2

* Starvation condition applies only for Experiment 5.

Table B.3. ORFs detected in multiple experiments

ORF	Gene	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7
YBR139W	YBR139W	-	-	UN	-	-	UN	-
YBR208C	DUR1,2	-	-	UN	-	-	UN	-
YDL239C	YDL239C	DR1-UR2	-	UN	-	-	UN	DR1-UR2
YDR090C	YDR090C	-	-	UN	-	-	UN	-
YGR190C	YGR190C	-	-	UN	-	-	UN	-
YHL016C	DUR3	DR1-UR2	-	UN	-	UN	UN	DR1-UR2
YHR029C	YHR029C	-	-	UN	-	-	UN	-
YHR037W	PUT2	-	-	UN	-	-	UN	-
YIL089W	YIL089W	DR1-UR2	-	UN	-	-	UN	DR1-UR2
YIL168W	SDL1	-	-	UN	-	-	UN	-
YIR027C	DAL1	-	-	UN	-	-	UN	-
YIR028W	DAL4	-	-	UN	-	UN	UN	-
YIR029W	DAL2	-	-	UN	-	-	UN	-
YIR030C	DCG1	-	-	UN	-	-	UN	-
YIR031C	DAL7	-	-	UN	-	-	UN	-
YJL172W	CPS1	-	-	UN	-	-	UN	-
YJR152W	DAL5	-	-	UN	-	UN	UN	-
YKR034W	DAL80	-	-	UN	-	-	UN	-
YKR039W	GAP1	-	-	UN	-	-	UN	-
YLR053C	YLR053C	-	-	UN	-	UN	UN	-
YLR142W	PUT1	-	-	UN	-	-	UN	-
YLR299W	ECM38	DR1-UR2	-	UN	-	-	UN	DR1-UR2
YMR088C	YMR088C	-	-	UN	-	-	UN	-
YMR090W	YMR090W	-	-	UN	-	-	UN	-
YMR170C	ALD2	-	-	UN	DH	-	UN	-
YNL117W	MLS1	-	-	UN	DH	-	UN	-
YNL142W	MEP2	-	-	UN	-	-	UN	-
YOL128C	YOL128C	-	-	UN	-	-	UN	-
YOR313C	SPS4	-	-	UN	-	-	UN	-
YOR348C	PUT4	-	-	UN	-	-	UN	-
YPR194c	opt2	-	-	UN	-	-	UN	-
YIR032C	DAL3	DH	-	UN	-	-	UN	-
YBR054W	YRO2	-	UN-DC	-	-	UN	-	-
YHR094C	HXT1	UN-DC	UN-DC	DC	-	-	DC	-
YDR345C	HXT3	UN-DC	-	DC	-	-	DC	-
YER028C	YER028C	UN-DC	-	DC	-	-	-	-

Table B.3. – continued

YGL157W	YGL157W	UN-DC	-	DC	-	-	DC	-
YGR138C	TPO2	-	UN-DC	DC	-	-	DC	-
YCR084C	TUP1	-	-	DC	-	-	DC	-
YER188W	YER188W	-	-	DC	-	-	DC	-
YGL209W	MIG2	-	-	DC	-	-	DC	-
YHL028W	WSC4	-	-	DC	-	-	DC	-
YHR092C	HXT4	-	-	DC	-	-	DC	-
YLR452C	SST2	-	-	DC	-	-	DC	-
YOR047C	STD1	-	-	DC	-	-	DC	-
YPL245W	YPL245W	-	-	DC	-	-	DC	-
YPR156C	TPO3	-	-	DC	-	-	DC	-
YDR508C	GNP1	-	-	DN	-	-	DN	-
YFR053C	HXK1	DN-UC	DN-UC	UC	-	-	UC	-
YGR243W	YGR243W	DN-UC	DN-UC	UC	-	UN	UC	-
YLR327C	YLR327C	DN-UC	DN-UC	UC	-	-	UC	-
YMR081C	ISF1	DN-UC	DN-UC	UC	-	-	UC	-
YOR374W	ALD4	DN-UC	DN-UC	UC	-	-	UC	-
YCL040W	GLK1	DN-UC	DN-UC	-	-	-	-	-
YDR342C	HXT7	DN-UC	DN-UC	-	-	-	-	-
YDR343C	HXT6	DN-UC	DN-UC	-	-	-	-	-
YGR174C	CBP4	DN-UC and UR1-DR2	DN-UC	-	-	-	-	UR1-DR2
YBR299W	MAL32	DN-UC	-	UC	-	-	UC	-
YDL079C	MRK1	DN-UC	-	UC	-	-	UC	-
YDR216W	ADR1	DN-UC	-	UC	-	-	-	-
YER121W	YER121W	DN-UC	-	UC	-	-	-	-
YFR017C	YFR017C	DN-UC	-	UC	-	-	UC	-
YGR067C	YGR067C	DN-UC	-	UC	-	-	-	-
YGR289C	MAL11	DN-UC	-	UC	-	-	UC	-
YIL155C	GUT2	DN-UC	-	UC	-	-	-	-
YKL093W	MBR1	DN-UC	-	UC	DH	-	-	-
YKL187C	YKL187C	DN-UC	-	UC	-	-	-	-
YKL217W	JEN1	DN-UC	-	UC	-	UC	UC	-
YBR068C	BAP2	DN-UC	-	DN-UC	-	-	-	-
SUC4	SUC4	-	-	UC	-	-	UC	-

Table B.3. – continued

YAL054C	ACS1	-	-	UC	DH	UN and UC	-	-
YCR010C	YCR010C	-	-	UC	-	UC	-	-
YDR256C	CTA1	-	-	UC	-	UC	-	-
YGR236C	YGR236C	-	-	UC	-	UN and UC	-	-
YHR096C	HXT5	-	-	UC	-	UN and UC	-	-
YIL057C	YIL057C	-	-	UC	-	UN	UC	-
YIL160C	POT1	-	-	UC	-	UC	-	-
YIL162W	SUC2	-	-	UC	-	-	UC	-
YMR096W	SNZ1	-	-	UC	-	UC	-	-
YMR107W	YMR107W	-	-	UC	-	UN and UC	-	-
YMR206W	YMR206W	-	-	UC	-	-	UC	-
YNL009W	IDP3	-	-	UC	-	UC	-	-
YNL194C	YNL194C	-	-	UC	-	UC	-	-
YNL202W	SPS19	-	-	UC	-	UC	-	-
YNR002C	FUN34	-	-	UC	-	UC	-	-
YNR034W	SOL1	-	-	UC	-	-	UC	-
YOR100C	CRC1	-	-	UC	-	UN and UC	-	-
YPL147W	PXA1	-	-	UC	-	UC	-	-
YPL248C	GAL4	-	-	UC	-	-	UC	-
YPR002W	YPR002W	-	-	UC	-	-	UC	-
YPR030W	YPR030W	-	-	UC	-	-	UC	-
YER067W	YER067W	-	DN-UC	UC	-	-	UC	-
YBL045C	COR1	DH	DN-UC	-	DH	-	-	-
YBL030C	PET9	DH	-	-	DH	-	-	-
YBR039W	ATP3	DH	-	-	DH	-	-	-
YDL004W	ATP16	DH	-	-	DH	-	-	-
YDR298C	ATP5	DH	-	-	DH	-	-	-

Table B.3. – continued

YDR322C-A	TIM11	DH	-	-	DH	-	-	-
YDR377W	ATP17	DH	-	-	DH	-	-	-
YDR529C	QCR7	DH	-	-	DH	-	-	-
YEL024W	RIP1	DH	-	-	DH	-	-	-
YGL187C	COX4	DH	-	-	DH	-	-	-
YGL191W	COX13	DH	-	-	DH	-	-	-
YGR183C	QCR9	DH	-	-	DH	-	-	-
YHR001W-A	QCR10	DH	-	-	DH	-	-	-
YHR051W	COX6	DH	-	-	DH	-	-	-
YJL102W	MEF2	DH	-	-	DH	-	-	-
YJL103C	YJL103C	DH	-	-	DH	-	-	-
YKL016C	ATP7	DH	-	-	DH	-	-	-
YLL001W	DNM1	DH	-	-	DH	-	-	-
YLR038C	COX12	DH	-	-	DH	-	-	-
YLR395C	COX8	DH	-	-	DH	-	-	-
YML081C-A	ATP18	DH	-	-	DH	-	-	-
YMR256C	COX7	DH	-	-	DH	-	-	-
YNL052W	COX5A	DH	-	-	DH	-	-	-
YNL284C	MRPL10	DH	-	-	DH	-	-	-
YNL315C	ATP11	DH	-	-	DH	-	-	-
YOR187W	TUF1	DH	-	-	DH	-	-	-
YPL271W	ATP15	DH	-	-	DH	-	-	-
YPR020W	ATP20	DH	-	-	DH	-	-	-
YPR191W	QCR2	DH	-	-	DH	-	-	-
YBR048W	RPS11B	UR1-DR2	-	-	-	-	-	UR1-DR2
YPL227C	ALG5	UR1-DR2	-	-	-	-	-	UR1-DR2
YDL100C	ARR4	UR1-DR2	-	-	-	-	-	UR1-DR2
YJR098C	YJR098C	UR1-DR2	-	-	-	-	-	UR1-DR2
YNL310C	YNL310C	UR1-DR2	-	-	-	-	-	UR1-DR2
YBR171W	SEC66	UR1-DR2	-	-	-	-	-	UR1-DR2
YNL108C	YNL108C	UR1-DR2	-	-	-	-	-	UR1-DR2
YKL097C	YKL097C	UR1-DR2	-	-	-	-	-	UR1-DR2
YDR023W	SES1	UR1-DR2	-	-	-	-	-	UR1-DR2
YER146W	LSM5	UR1-DR2	-	-	-	-	-	UR1-DR2
YEL056W	HAT2	UR1-DR2	-	-	-	-	-	UR1-DR2
YMR235C	RNA1	UR1-DR2	-	-	-	-	-	UR1-DR2
YGL076C	RPL7A	UR1-DR2	-	-	-	-	-	UR1-DR2

Table B.3. – continued

YDL105W	QRI2	UR1-DR2	-	-	-	-	-	UR1-DR2
YGL097W	SRM1	UR1-DR2	-	-	-	-	-	UR1-DR2
YKL036C	YKL036C	UR1-DR2	-	-	-	-	-	UR1-DR2
YHR062C	RPP1	UR1-DR2	-	-	-	-	-	UR1-DR2
YNL181W	YNL181W	UR1-DR2	-	-	-	-	-	UR1-DR2
YER006W	NUG1	UR1-DR2	-	-	-	-	-	UR1-DR2
YBL039C	URA7	UR1-DR2	-	-	-	-	-	UR1-DR2
YPR110C	RPC40	UR1-DR2	-	-	-	-	-	UR1-DR2
YOR092W	ECM3	UR1-DR2	-	-	-	-	-	UR1-DR2
YJR147W	HMS2	UR1-DR2	-	-	-	-	-	UR1-DR2
YCR034W	FEN1	UR1-DR2	-	-	-	-	-	UR1-DR2
YJR069C	HAM1	UR1-DR2	-	-	-	-	-	UR1-DR2
YLL034C	YLL034C	UR1-DR2	-	-	-	-	-	UR1-DR2
YDR144C	MKC7	UR1-DR2	-	-	-	-	-	UR1-DR2
YOR243C	PUS7	UR1-DR2	-	-	-	-	-	UR1-DR2
YGR123C	PPT1	UR1-DR2	-	-	-	-	-	UR1-DR2
YLR401C	YLR401C	UR1-DR2	-	-	-	-	-	UR1-DR2
YNL175C	NOP13	UR1-DR2	-	-	-	-	-	UR1-DR2
YPL266W	DIM1	UR1-DR2	-	-	-	-	-	UR1-DR2
YGR173W	YGR173W	UR1-DR2	-	-	-	-	-	UR1-DR2
YDR408C	ADE8	UR1-DR2	-	-	-	-	-	UR1-DR2
YKL143W	LTV1	UR1-DR2	-	-	-	-	-	UR1-DR2
YJL200C	YJL200C	UR1-DR2	-	-	-	-	-	UR1-DR2
YOR199W	YOR199W	UR1-DR2	-	-	-	-	-	UR1-DR2
YFR005C	SAD1	UR1-DR2	-	-	-	-	-	UR1-DR2
YGL121C	GPG1	DR1-UR2	-	-	-	-	-	DR1-UR2
YDR070C	YDR070C	DR1-UR2	-	-	-	-	-	DR1-UR2
YBL048W	YBL048W	DR1-UR2	-	-	-	-	-	DR1-UR2
YGR088W	CTT1	DR1-UR2	-	-	-	-	-	DR1-UR2
YOL084W	PHM7	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR173W	YOR173W	DR1-UR2	-	-	-	-	-	DR1-UR2
YMR250W	GAD1	DR1-UR2	-	-	-	-	-	DR1-UR2
YMR181C	YMR181C	DR1-UR2	-	-	-	-	-	DR1-UR2
YOL153C	YOL153C	DR1-UR2	-	-	-	-	-	DR1-UR2
YGR066C	YGR066C	DR1-UR2	-	-	-	-	-	DR1-UR2
YPR077C	YPR077C	DR1-UR2	-	-	-	-	-	DR1-UR2
YLL020C	YLL020C	DR1-UR2	-	-	-	-	-	DR1-UR2

Table B.3. – continued

YGL156W	AMS1	DR1-UR2	-	-	-	-	-	DR1-UR2
YMR173W	DDR48	DR1-UR2	-	-	-	-	-	DR1-UR2
YLL020C	YLL020C	DR1-UR2	-	-	-	-	-	DR1-UR2
YIL017C	VID28	DR1-UR2	-	-	-	-	-	DR1-UR2
YGR008C	STF2	DR1-UR2	-	-	-	-	-	DR1-UR2
YGL166W	CUP2	DR1-UR2	-	-	-	-	-	DR1-UR2
YDR001C	NTH1	DR1-UR2	-	-	-	-	-	DR1-UR2
YGL096W	TOS8	DR1-UR2	-	-	-	-	-	DR1-UR2
YEL011W	GLC3	DR1-UR2	-	-	-	-	-	DR1-UR2
YGR070W	ROM1	DR1-UR2	-	-	-	-	-	DR1-UR2
YLR431C	YLR431C	DR1-UR2	-	-	-	-	-	DR1-UR2
YPR026W	ATH1	DR1-UR2	-	-	-	-	-	DR1-UR2
YJR161C	COS5	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR031W	CRS5	DR1-UR2	-	-	-	-	-	DR1-UR2
YIL097W	FYV10	DR1-UR2	-	-	-	-	-	DR1-UR2
YKL188C	PXA2	DR1-UR2	-	-	-	-	-	DR1-UR2
YLR311C	YLR311C	DR1-UR2	-	-	-	-	-	DR1-UR2
YDR313C	PIB1	DR1-UR2	-	-	-	-	-	DR1-UR2
YJR155W	AAD10	DR1-UR2	-	-	-	-	-	DR1-UR2
YLR457C	NBP1	DR1-UR2	-	-	-	-	-	DR1-UR2
YHR171W	APG7	DR1-UR2	-	-	-	-	-	DR1-UR2
YLR271W	YLR271W	DR1-UR2	-	-	-	-	-	DR1-UR2
YDR319C	YDR319C	DR1-UR2	-	-	-	-	-	DR1-UR2
YDR525W-A	SNA2	DR1-UR2	-	-	-	-	-	DR1-UR2
YBR169C	SSE2	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR161C	YOR161C	DR1-UR2	-	-	-	-	-	DR1-UR2
YDR208W	MSS4	DR1-UR2	-	-	-	-	-	DR1-UR2
YIL077C	YIL077C	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR366W	YOR366W	DR1-UR2	-	-	-	-	-	DR1-UR2
YCL052C	PBN1	DR1-UR2	-	-	-	-	-	DR1-UR2
YEL023C	YEL023C	DR1-UR2	-	-	-	-	-	DR1-UR2
YDL138W	RGT2	DR1-UR2	-	-	-	-	-	DR1-UR2
YPL274W	SAM3	DR1-UR2	-	-	-	-	-	DR1-UR2
YNL192W	CHS1	DR1-UR2	-	-	-	-	-	DR1-UR2
YMR305C	SCW10	DR1-UR2	-	-	-	-	-	DR1-UR2
YNL156C	YNL156C	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR296W	YOR296W	DR1-UR2	-	-	-	-	-	DR1-UR2

Table B.3. – continued

YNL225C	CNM67	DR1-UR2	-	-	-	-	-	DR1-UR2
YPL082C	MOT1	DR1-UR2	-	-	-	-	-	DR1-UR2
YLR396C	VPS33	DR1-UR2	-	-	-	-	-	DR1-UR2
YBL017C	PEP1	DR1-UR2	-	-	-	-	-	DR1-UR2
YIL159W	BNR1	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR285W	YOR285W	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR219C	STE13	DR1-UR2	-	-	-	-	-	DR1-UR2
YNL020C	ARK1	DR1-UR2	-	-	-	-	-	DR1-UR2
YJR055W	HIT1	DR1-UR2	-	-	-	-	-	DR1-UR2
YPL020C	ULP1	DR1-UR2	-	-	-	-	-	DR1-UR2
YDR507C	GIN4	DR1-UR2	-	-	-	-	-	DR1-UR2
YKL048C	ELM1	DR1-UR2	-	-	-	-	-	DR1-UR2
YLR067C	PET309	DR1-UR2	-	-	-	-	-	DR1-UR2
YFL034W	YFL034W	DR1-UR2	-	-	-	-	-	DR1-UR2
YBR010W	HHT1	DR1-UR2	-	-	-	-	-	DR1-UR2
YOR155C	ISN1	DR1-UR2	-	-	-	-	-	DR1-UR2
YNR059W	MNT4	DR1-UR2	-	-	-	-	-	DR1-UR2
YMR172W	HOT1	DR1-UR2	-	-	-	-	-	DR1-UR2
YJL166W	QCR8	DR1-UR2	-	-	-	-	-	DR1-UR2
YLR324W	YLR324W	DR1-UR2	-	-	-	-	-	DR1-UR2

Table B.4. Results for the mapping of biological process GO terms on genes affected by conditions UN-CD, DN-UC and DH

GOID	GO term	Frequency (/7274)	UN-DC (/47)	DN-UC (/45)	DH (/29)	P_{UN-DC}	P_{DN-UC}	P_{DH}	ORF_{UN-DC}	ORF_{DN-UC}	ORF_{DH}
4	Process: biological process unknown	2468	11	12	0	9.57E-01	8.84E-01	1.00E+00	YBR139W YDR090C YGR190C YHR029C YIL089W YLR053C YMR090W YRO2 YGL157W YER188W YPL245W	FMP43 RBF9 YER121W YFR017C YGR067C YKL187C SPG1 YIL057C SPG4 YMR206W YNL194C YER067W	none
746	Process: conjugation	100	1	0	0	4.78E-01	1.00E+00	1.00E+00	SST2	none	none
910	Process: cytokinesis	122	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none

Table B.4 – continued

5975	Process: carbohydrate metabolism	219	3	9	0	1.68E-01	6.73E-06	1.00E+00	MLS1 MIG2 STD1	HXK1 GLK1 MAL32 ADR1 GUT2 SUC4 SUC2 IDP3 GAL4	none
6091	Process: generation of precursor metabolites and energy	226	1	5	23	7.73E-01	1.26E-02	8.35E-30	MLS1	ISF1 ADR1 MBR1 ACS1 IDP3	COR1 PET9 ATP3 ATP16 ATP5 TIM11 ATP17 QCR7 RIP1 COX4 COX13 QCR9 QCR10 COX6 YJL103C ATP7 COX8 ATP18 COX7 COX5A ATP15 ATP20 QCR2
6118	Process: electron transport	33	0	0	10	1.00E+00	1.00E+00	6.84E-17	none	none	QCR7 RIP1 COX4 QCR9 QCR10 COX6 COX8 COX7 COX5A QCR2

Table B.4 – continued

6259	Process: DNA metabolism	554	2	1	0	8.82E-01	9.72E-01	1.00E+00	SPS4 TUP1	ADY2	none
6350	Process: transcription	491	5	3	0	2.09E-01	5.93E-01	1.00E+00	DAL80 MIG3 TUP1 MIG2 STD1	ADR1 GAL4 CSR2	none
6412	Process: protein biosynthesis	805	0	0	3	1.00E+00	1.00E+00	6.37E-01	none	none	MEF2 MRPL10 TUF1
6464	Process: protein modification	429	1	1	0	9.43E-01	9.35E-01	1.00E+00	YGK3	MRK1	none
6519	Process: amino acid and derivative metabolism	195	4	0	0	3.70E-02	1.00E+00	1.00E+00	PUT2 PUT1 ALD2 PUT4	none	none
6629	Process: lipid metabolism	216	0	4	0	1.00E+00	4.42E-02	1.00E+00	none	POT1 IDP3 SPS19 CRC1	none
6766	Process: vitamin metabolism	72	0	3	0	1.00E+00	1.01E-02	1.00E+00	none	GUT2 SNZ1 IDP3	none

Table B.4 – continued

6810	Process: transport	899	13	11	10	3.71E-03	1.89E-02	1.77E-03	DUR3 DAL4 DAL5 GAP1 VBA1 MEP2 PUT4 OPT2 HXT1 HXT3 TPO2 HXT4 TPO3	GNP1 HXT7 HXT6 MAL11 JEN1 BAP2 ADY2 HXT5 FUN34 SOL1 PXA1	PET9 ATP3 ATP16 ATP5 TIM11 ATP17 ATP7 ATP18 ATP15 ATP20
6950	Process: response to stress	397	4	1	0	2.54E-01	9.20E-01	1.00E+00	YGK3 MIG3 WSC4 STD1	MRK1	none
6996	Process: organelle organization and biogenesis	1009	2	1	1	9.92E-01	9.99E-01	9.87E-01	TUP1 WSC4	ADR1	DNM1
6997	Process: nuclear organization and biogenesis	64	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none
7010	Process: cytoskeleton organization and biogenesis	294	1	0	0	8.56E-01	1.00E+00	1.00E+00	WSC4	none	none
7047	Process: cell wall organization and biogenesis	147	1	1	0	6.17E-01	6.01E-01	1.00E+00	WSC4	CSR2	none

Table B.4 – continued

7049	Process: cell cycle	407	1	1	0	9.33E-01	9.25E-01	1.00E+00	SPS4	ADY2	none
7114	Process: cell budding	77	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none
7124	Process: pseudohyphal growth	48	1	0	0	2.67E-01	1.00E+00	1.00E+00	MEP2	none	none
7126	Process: meiosis	133	1	1	0	5.80E-01	5.64E-01	1.00E+00	SPS4	ADY2	none
7165	Process: signal transduction	169	3	0	0	9.58E-02	1.00E+00	1.00E+00	WSC4 SST2 STD1	none	none
9653	Process: morphogenesis	147	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none
16044	Process: membrane organization and biogenesis	0	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none
16070	Process: RNA metabolism	537	0	1	0	1.00E+00	9.68E-01	1.00E+00	none	SOL1	none
16192	Process: vesicle-mediated transport	299	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none
19725	Process: cell homeostasis	109	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none
30163	Process: protein catabolism	176	2	1	0	3.15E-01	6.68E-01	1.00E+00	CPS1 YGK3	MRK1	none

Table B.4 – continued

30435	Process: sporulation	102	2	1	0	1.41E-01	4.70E-01	1.00E+00	ADY3 SPS4	SPS19	none
42254	Process: ribosome biogenesis and assembly	254	0	0	0	1.00E+00	1.00E+00	1.00E+00	none	none	none
45333	Process: cellular respiration	88	0	2	13	1.00E+00	1.03E-01	6.74E-18	none	ISF1 MBR1	COR1 PET9 QCR7 RIP1 COX4 COX13 QCR9 QCR10 COX6 COX8 COX7 COX5A QCR2
	not_yet_annotated	-	1	0	0	-	-	-	YIL168W		
other	other	-	7	4	2	-	-	-	DUR1,2 DAL1 DAL2 DCG1 DAL7 ECM38 DAL3	ALD4 CBP4 CTA1 PDH1	COX12 ATP11

Table B.5. Results for the mapping of biological process GO terms on ORFs respond to increase in dilution rate

GOID	GO term	Frequency (/7274)	UR2-DR1(/36)	DR2-UR1 (/47)	$P_{UR2-DR1}$	$P_{DR2-UR1}$	$ORFs_{UR2-DR1}$	$ORFs_{DR2-UR1}$
4	Process: biological process unknown	2468	7	20	9.82E-01	1.37E-01	YJR098C YKL097C YKL036C YNL181W NOP13 RBG2 YJL200C	FMP16 YBL048W YMR090W PHM7 DCS2 YMR181C YOL153C YGR066C YPR077C YLL020C COS5 YLR311C NBP1 YLR271W YDR319C SNA2 PNS1 YIL077C YOR366W YEL023C
746	Process: conjugation	100	0	0	1.00E+00	1.00E+00	none	none
910	Process: cytokinesis	122	0	0	1.00E+00	1.00E+00	none	none
5975	Process: carbohydrate metabolism	219	0	6	1.00E+00	2.77E-03	none	AMS1 VID28 NTH1 GLC3 ATH1 FYV10
6091	Process: generation of precursor metabolites and energy	226	0	8	1.00E+00	9.23E-05	none	ISF1 VID28 STF2 MBR1 NTH1 GLC3 ATH1 FYV10

Table B.5. - continued

6118	Process: electron transport	33	0	0	1.00E+00	1.00E+00	none	none
6259	Process: DNA metabolism	554	3	1	5.24E-01	9.76E-01	HAT2 NSE4 HAM1	DDR48
6350	Process: transcription	491	2	2	7.09E-01	8.35E-01	HAT2 RPC40	CUP2 TOS8
6412	Process: protein biosynthesis	805	4	1	5.75E-01	9.96E-01	RPS11B ALG5 SES1 RPL7A	PBN1
6464	Process: protein modification	429	2	2	6.35E-01	7.73E-01	ALG5 PPT1	PIB1 PBN1
6519	Process: amino acid and derivative metabolism	195	1	1	6.24E-01	7.21E-01	SES1	GAD1
6629	Process: lipid metabolism	216	2	1	2.90E-01	7.58E-01	URA7 FEN1	PBN1
6766	Process: vitamin metabolism	72	0	0	1.00E+00	1.00E+00	none	none
6810	Process: transport	899	6	5	2.82E-01	7.06E-01	ZIM17 SEC66 RNA1 SRM1 FEN1 RIX7	HXT5 STF2 ATG23 PXA2 ATG7
6950	Process: response to stress	397	4	6	1.31E-01	4.17E-02	ARR4 NSE4 HAM1 LTV1	CTT1 GAD1 DDR48 STF2 NTH1 ATH1

Table B.5 – continued

6996	Process: organelle organization and biogenesis	1009	11	2	7.73E-03	9.92E-01	RPS11B ZIM17 LSM5 HAT2 RNA1 SRM1 RPP1 NUG1 RIX7 DIM1 LTV1	ROM1 MSS4
6997	Process: nuclear organization and biogenesis	64	0	0	1.00E+00	1.00E+00	none	none
7010	Process: cytoskeleton organization and biogenesis	294	0	2	1.00E+00	5.71E-01	none	ROM1 MSS4
7047	Process: cell wall organization and biogenesis	147	1	1	5.20E-01	6.17E-01	ECM3	ROM1
7049	Process: cell cycle	407	0	1	1.00E+00	9.33E-01	none	TOS8
7114	Process: cell budding	77	0	1	1.00E+00	3.94E-01	none	ROM1
7124	Process: pseudohyphal growth	48	1	0	2.12E-01	1.00E+00	HMS2	none

Table B.5 – continued

7126	Process: meiosis	133	0	0	1.00E+00	1.00E+00	none	none
7165	Process: signal transduction	169	0	3	1.00E+00	9.58E-02	none	GPG1 ROM1 RGT2
9653	Process: morphogenesis	147	0	1	1.00E+00	6.17E-01	none	ROM1
16044	Process: membrane organization and biogenesis	0	0	0	1.00E+00	1.00E+00	none	none
16070	Process: RNA metabolism	537	8	0	4.10E-03	1.00E+00	SES1 LSM5 RNA1 RPP1 NUG1 PUS7 DUS3 DIM1	none
16192	Process: vesicle-mediated transport	299	1	0	7.79E-01	1.00E+00	FEN1	none
19725	Process: cell homeostasis	109	0	0	1.00E+00	1.00E+00	none	none
30163	Process: protein catabolism	176	1	0	5.86E-01	1.00E+00	MKC7	none
30435	Process: sporulation	102	0	1	1.00E+00	4.85E-01	none	MSS4

Table B.5 – continued

42254	Process: ribosome biogenesis and assembly	254	9	0	3.08E-06	1.00E+00	RPS11B LSM5 RNA1 SRM1 RPP1 NUG1 RIX7 DIM1 LTV1	none
45333	Process: cellular respiration	88	0	2	1.00E+00	1.11E-01	none	ISF1 MBR1
other	other	-	2	3			YNL108C ADE8	CRS5 AAD10 SSE2

In the first column of Table B.6, the biological process terms common among the correlated ORFs are given. In the second column, the number of correlated ORFs annotated to the terms is given with the total number of genes correlated to unknown ORF. In the third column, total number of ORFs annotated to the terms is given with the total number of ORFs annotated in SGD. In the fourth column, significance of the term is given, based on the statistics given in columns two and three, the p-value is calculated by SGD using Eq.3.25. In the last column, the correlated ORFs annotated to the term are listed. Table B.6 is given in TableB6.doc file in the attached CD.

APPENDIX C: ALGORITHM OF PREDPOWER

In Figure C.1, two networks with several nodes and edges are given. The algorithm analyses each node in Network 1 assuming that it has no relation to Network 2 (which is equivalent to a gene with no annotation). The nodes of Network are given in parentheses "( )" and nodes of Network 2 are given in brackets "[ ]" in the following paragraphs.

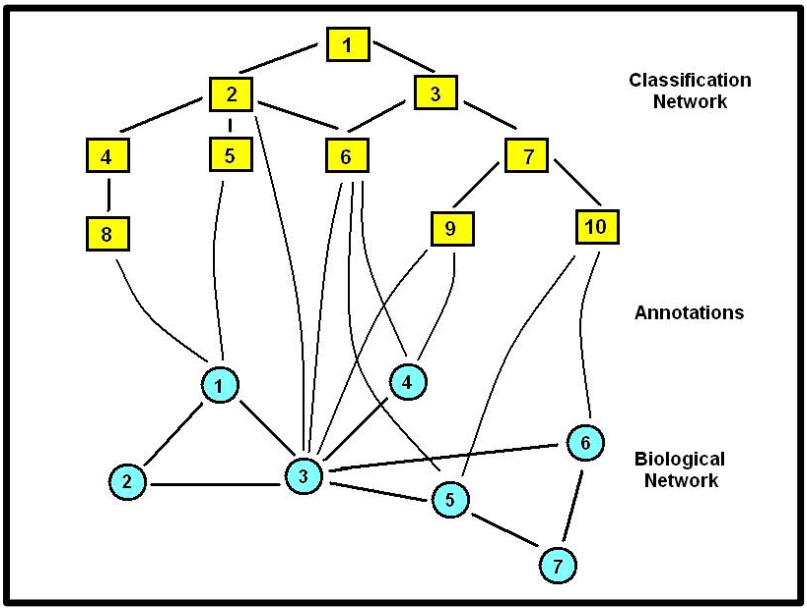

Figure C.1. General structure of networks processed by PredPower

First step of the algorithm is making a list of relations for each node from Network 1 to Network 2. The relations of the node itself is omitted (assuming that it has no relations) and only the relations from its neigbors to Network 2 are taken into consideration. (3) and (2) are the neighbors of (1) in Network1, and they are related to [3], [6], and [9]. Thus, the relation list of node (1) consists of three relations to [3], [6], and [9]. Table C.1 is constructed for all nodes from Network 1.

For the given two networks, N = 7, the other parameters are given on Tables C.2 and C.3. These values are used to calculate the p-values on Table C.4 as in Equation C.1. In Table C.4, the new relations are listed and compared to the existing ones.:

$$p = \sum_{j=x}^{n} \left(\frac{n!}{j!(n-j)!}\right) \phi^{j} (1-\phi)^{(n-j)} \tag{C.1}$$

where $\phi = G/N$

G: Number of all nodes from Network1 related to a specific node in Network2.

N: Number of all nodes in Network1

n: Number of neighbors of a node in Network1

x: Number of neighbors of a node from Network1 that has relation to a specific node in Network2.

Table C.1. List of Relations for Nodes of Network1 to Network2

Network1	Network2		Network1	Network2
(1)	[2]		(4)	[2]
(1)	[6]		(4)	[6]
(1)	[9]		(4)	[9]
(2)	[8]		(5)	[2]
(2)	[5]		(5)	[6]
(2)	[2]		(5)	[9]
(2)	[6]		(6)	[2]
(2)	[9]		(6)	[6]
(3)	[6] x 2		(6)	[9]
(3)	[8]		(7)	[6]
(3)	[9]		(7)	[10] x 2
(3)	[10] x 2			

Table C.2. Values of parameters “n” and “G”

Nodes	n		Nodes	G
(1)	2		[1]	0
(2)	2		[2]	1
(3)	5		[3]	0
(4)	1		[4]	0
(5)	2		[5]	1
(6)	2		[6]	3
(7)	2		[7]	0
			[8]	1
			[9]	2
			[10]	2

Table C.3. Values of parameter "x"

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
(1)	0	1	0	0	0	1	0	0	1	0
(2)	0	1	0	0	1	1	0	1	1	0
(3)	0	0	0	0	0	2	0	1	1	2
(4)	0	1	0	0	0	1	0	0	1	0
(5)	0	1	0	0	0	1	0	0	1	0
(6)	0	1	0	0	0	1	0	0	1	0
(7)	0	0	0	0	0	1	0	0	0	2

Table C.4. P-value

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
(1)	1.00	0.27	1.00	1.00	1.00	0.67	1.00	1.00	0.49	1.00
(2)	1.00	0.27	1.00	1.00	0.27	0.67	1.00	0.27	0.49	1.00
(3)	1.00	1.00	1.00	1.00	1.00	0.71	1.00	0.54	0.81	0.44
(4)	1.00	0.14	1.00	1.00	1.00	0.43	1.00	1.00	0.29	1.00
(5)	1.00	0.27	1.00	1.00	1.00	0.67	1.00	1.00	0.49	1.00
(6)	1.00	0.27	1.00	1.00	1.00	0.67	1.00	1.00	0.49	1.00
(7)	1.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0.08

Table C.5. Comparison of new relations with the existing relations

New Relation		Comparison	
Network1	Network2	p-value	Correct?
(1)	[2]	0.27	No
(1)	[6]	0.67	No
(1)	[9]	0.49	No
(2)	[2]	0.27	Novel
(2)	[5]	0.27	Novel
(2)	[6]	0.67	Novel
(2)	[8]	0.27	Novel
(2)	[9]	0.49	Novel
(3)	[6]	0.71	Yes
(3)	[8]	0.54	No
(3)	[9]	0.81	Yes
(3)	[10]	0.44	No
(4)	[2]	0.14	No
(4)	[6]	0.43	Yes
(4)	[9]	0.29	Yes
(5)	[2]	0.27	No
(5)	[6]	0.67	Yes
(5)	[9]	0.49	No
(6)	[2]	0.27	No
(6)	[6]	0.67	No
(6)	[9]	0.49	No
(7)	[6]	0.67	Novel
(7)	[10]	0.08	Novel

Table C.6. Evaluation of results ($p < 0.99$)

Fraction of known nodes with an estimated relation	5/5	1.00
Fraction of correctly estimated relations	5/16	0.31
Fraction of known nodes with a correctly estimate relation	3/5	0.6

Table C.7. Evaluation of results ($p < 0.50$)

Fraction of known nodes with an estimated relation	5/5	1.00
Fraction of correctly estimated relations	2/16	0.13
Fraction of known nodes with a correctly estimate relation	1/5	0.20

The steps given above considers only the direct annotations. The same procedure applies when neighboring terms are needed to be included. Table C.1 is modified to cover the neighboring terms and all steps are repeated. List of modified relations for Node (1) with first neighbors included case is given in Table C.8.

Table C.8. List of relations for nodes of Network1 to Network2

Network1	Network2
(1)	[1]
(1)	[2] x2
(1)	[3]
(1)	[6]
(1)	[7]
(1)	[9]

APPENDIX D: NOVEL ANNOTATIONS VIA PREDPOWER

The novel annotations made for unknown ORFs are presented in Tables D.1 and D.2. In Table D.1, GO biological process terms are assigned to ORFs with unknown biological process and in Table D.2. GO molecular function terms are assigned to ORFs with unknown molecular function. Only first few rows of Table D.1 is presented here, please refer to attached CD for the complete tables saved with the name AppendixD.doc.

ORFs are assigned the terms “biological process unknown” or “molecular function unknown” if the ORFs with similarity to those ORFs are also unknowns. These type of annotations will be removed from PredPower in future versions.

Table D.1. Biological Process Terms Annotated to Unkown ORFs by PredPower

ORF	Term	ORF	Term	P-value
34	7046	RBG1	ribosome biogenesis	1.2E-10
34	16070	RBG1	RNA metabolism	7.8E-10
34	7028	RBG1	cytoplasm organization and biogenesis	1.2E-09
34	42254	RBG1	ribosome biogenesis and assembly	1.2E-09
34	6364	RBG1	rRNA processing	1.6E-08
57	5975	YAL061W	carbohydrate metabolism	5.3E-15
57	44262	YAL061W	cellular carbohydrate metabolism	7.1E-14
57	6091	YAL061W	generation of precursor metabolites and energy	9.4E-13
57	15980	YAL061W	energy derivation by oxidation of organic compounds	5.1E-11
57	6066	YAL061W	alcohol metabolism	6.2E-11
57	50896	YAL061W	response to stimulus	1.6E-09
57	5996	YAL061W	monosaccharide metabolism	3.3E-09
57	6950	YAL061W	response to stress	1.2E-08
60	4	YAL064W	biological process unknown	5.0E-08
73	8202	PAU7	steroid metabolism	4.1E-12
73	6694	PAU7	steroid biosynthesis	1.2E-11
73	16125	PAU7	sterol metabolism	1.6E-11
73	16126	PAU7	sterol biosynthesis	5.2E-11
73	6696	PAU7	ergosterol biosynthesis	1.6E-08
73	8204	PAU7	ergosterol metabolism	1.6E-08
88	128	YAR062W	flocculation	1.9E-08
88	501	YAR062W	flocculation (sensu Saccharomyces)	1.9E-08
88	16339	YAR062W	calcium-dependent cell-cell adhesion	1.9E-08
88	7157	YAR062W	heterophilic cell adhesion	8.3E-08

APPENDIX E: MATLAB CODE OF PREDPOWER

The MATLAB code of PredPower is given in the attached CD as .m files under the directory "AppendixE". Sample networks are also available to test the software. Please refer to "readme.txt" file before running the programs.

REFERENCES

Affymetrix, 2000-2003, *Affymetrix GeneChip Expression Analysis Technical Manual*, Affymetrix Inc., USA.

Akache B., K. Wu and B. Turcotte , 2001, "Phenotypic analysis of genes encoding yeast zinc cluster proteins", *Nucleic Acids Research*, Vol. 29, No. 10, pp. 2181-2190.

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

Alfarano C., C. E. Andrade, K. Anthony, N. Bahroos, M. Bajec, K. Bantoft, D. Betel, B. Bobechko, K. Boutilier, E. Burgess, K. Buzadzija, R. Cavero, C. D'Abreo, I. Donaldson, D. Dorairajoo, M. J. Dumontier, M. R. Dumontier, V. Earles, R. Farrall, H. Feldman, E. Garderman, Y. Gong, R. Gonzaga, V. Grytsan, E. Gryz, V. Gu, E. Haldorsen, A. Halupa, R. Haw, A. Hrvojic, L. Hurrell, R. Isserlin, F. Jack, F. Juma, A. Khan, T. Kon, S. Konopinsky, V. Le, E. Lee, S. Ling, M. Magidin, J. Moniakis, J. Montojo, S. Moore, B. Muskat, I. Ng, J.P. Paraiso, B. Parker, G. Pintilie, R. Pirone, J.J. Salama, S. Sgro, T. Shan, Y. Shu, J. Siew, D. Skinner, K. Snyder, R. Stasiuk, D. Strumpf, B. Tuekam, S. Tao, Z. Wang, M. White, R. Willis, C. Wolting, S. Wong, A. Wrong, C. Xin, R. Yao, B. Yates, S. Zhang, K. Zheng, T. Pawson, B. F. Ouellette and C. W. Hogue, 2005, "The Biomolecular Interaction Network Database and related tools 2005 update", *Nucleic Acids Research*, Vol. 33, Database Issue, pp. D418-D424.

Allen, J., H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver and D. B. Kell, 2003, "High-throughput classification of yeast mutants for functional genomics using metabolic footprinting", *Nature Biotechnology*, Vol. 21, No. 6, pp. 692-696, June.

Allocco D. J., I. S. Kohane and A. J. Butte, 2004, "Quantifying the relationship between co-expression, co-regulation and gene function", *BMC Bioinformatics*, Vol.5, No.1, pp. 18-24.

Altamura N., N. Capitanio, N. Bonnefoy, S. Papa and G. Dujardin, 1996, "The *Saccharomyces cerevisiae OXA1* gene is required for the correct assembly of cytochrome c oxidase and oligomycin-sensitive ATP synthase", *FEBS Letters*, Vol. 382, No. 1-2, pp. 111-115.

Baganz F., A. Hayes, D. Marren, D.C. Gardner and S.G. Oliver, 1997, "Suitability of replacement markers for functional analysis studies in *Saccharomyces cerevisiae*", *Yeast*, Vol. 16, No. 13, pp. 1563-1573.

Bailey J. E. and D. F. Ollis, 1986, *Biochemical Engineering Fundamentals*, Second Edition, McGraw-Hill, Singapore.

Baxter S. M., J. S. Rosenblum, S. Knutson, M. R. Nelson, J. S. Montimurro, J. A. Di Gennaro, J. A. Speir, J. J. Burbaum and J. S. Fetrow, 2004, "Synergistic computational and experimental proteomics approaches for more accurate detection of active serine hydrolases in yeast", *Molecular and Cellular Proteomics*, Vol. 3, No. 3, pp. 209-225, March.

Berden J. A., H. Boumans and L. Grivell, 1998, "The respiratory chain in yeast behaves as a single functional unit", *The Journal of Biological Chemistry*, Vol. 273, No. 9, pp. 4872-4877.

Blom, J., J. T. de Mattos, M. J. Teixeira, and L. A. Grivell, 2000, "Redirection of the Respiro-Fermentative Flux Distribution in *Saccharomyces cerevisiae* by Overexpression of the Transcription Factor HAP4p", *Applied and Environmental Microbiology*, Vol. 66, No. 5, pp. 1970–1973.

Boer, V. M., J. H. de Winde, J. T. Pronk, and M. D. W. Piper, 2003, "The Genome-wide Transcriptional Responses of *Saccharomyces cerevisiae* Grown on Glucose in Aerobic Chemostat Cultures Limited for Carbon, Nitrogen, Phosphorus, or Sulfur", *The Journal of Biochemistry*, Vol. 278, No. 5, pp. 3265-3274.

Bohm S., D. Frishman and H. W. Mewes, 1997, "Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins", *Nucleic Acids Research*, Vol. 25, No. 12, pp. 2464-2469.

Bott, M., and A. Niebisch, 2003, "The respiratory chain of *Corynebacterium glutamicum*", *Journal of Biotechnology*, Vol. 104, No. 1, pp. 129-153.

Brand, M. D., and R. K. Curtis, 2002, "Control analysis of DNA microarray expression data", *Molecular Biology Reports*, Vol. 29, No. 1, pp. 67-71.

Breitkreutz B. J., C. Stark and M. Tyers, 2003, "The GRID: the General Repository for Interaction Datasets", *Genome Biology*, Vol. 4, No. 3, pp. R23-R23.

Brewster N. K., D. L. Val, M. E. Walker and J. C. Wallace, 1994, "Regulation of pyruvate carboxylase isozyme (*PYC1*, *PYC2*) gene expression in *Saccharomyces cerevisiae* during fermentative and nonfermentative growth", *Archives of Biochemistry and Biophysics*, Vol. 311, No. 1, pp. 62-71.

Bro C., B. Regenberg, G. Lagniel, J. Labarre, M. Montero-Lomeli and J. Nielsen, 2003, "Transcriptional, proteomic, and metabolic responses to lithium in galactose-grown yeast cells", *Journal of Biological Chemistry*, Vol. 278, No. 34, pp. 32141-32149.

Brun C., C. Herrmann and A. Guenoche, 2004, "Clustering proteins from interaction networks for the prediction of cellular functions", *BMC Bioinformatics*, Vol. 5, pp. 95.

Buchholtz, A., J. Hurlebaus, C. Wandrey and R. Takors, 2002, "Metabolomics: quantification of intracellular metabolite dynamics", *Biomolecular Engineering*, Vol. 19, No. 1, pp. 5-15.

Buschlen S., J. M. Amillet, B. Guiard, A. Fournier, C. Marcireau and M. Bolotin-Fukuhara, 2003, "The *S. cerevisiae* HAP complex, a key regulator of mitochondrial

function, coordinates nuclear and mitochondrial gene expression", *Comparative and Functional Genomics*, Vol. 4, pp. 37–46.

Butow R. A. and N. G. Avadhani, 2004, "Mitochondrial signaling: the retrograde response", *Molecular Cell*, Vol. 14, No. 1, pp. 1-15.

Caba E., D. A. Dickinson, G. R. Warnes and J. Aubrecht, 2005, "Differentiating mechanisms of toxicity using global gene expression analysis in *Saccharomyces cerevisiae*", *Mutation Research*, Vol. 575, No. 1-2, pp. 34-46.

Castrillo J. I. and S. G. Oliver, 2004, "Yeast as a Touchstone in Post-genomic Research: Strategies for Integrative Analysis in Functional Genomics", *Journal of Biochemistry and Moleculer Biology*, Vol. 37, No. 1, pp. 93-106.

Chen Y. and D. Xu, 2004, "Global protein function annotation through mining genome-scale data in yeast *Saccharomyces cerevisiae*", *Nucleic Acids Research*, Vol. 32, No. 21, pp. 6414-6424.

Cho Y and V. Walbot, 2001, "Computational methods for gene annotation: the *Arabidopsis* genome", *Current Opinion in Biotechnology*, Vol. 12, pp. 126–130.

Cornell, M., N. W. Paton, C. Hedeler, P. Kirby, D. Delneri, A. Hayes and S. G. Oliver, 2003, "GIMS: an integrated data storage and analysis environment for genomic and functional data", *Yeast*, Vol. 20, No. 15, pp. 1291-306.

Crespo J. L. and M. N. Hall, 2002, "Elucidating TOR signaling and rapamycin action: lessons from *Saccharomyces cerevisiae*", *Microbiology and Molecular Biolology Reviews*, Vol. 66, No. 4, pp. 579-591.

Crivellone M. D., M. A. Wu and A. Tzagoloff, 1988, "Assembly of the mitochondrial membrane system. Analysis of structural mutants of the yeast coenzyme QH2-cytochrome c reductase complex", *Journal of Biological Chemistry*, Vol. 263, No. 28, pp. 14323-14333.

Curwen V., E. Eyrasi ,T. D. Andrews, L. Clarke, E. Mongin, S. M. Searle and M. Clamp, 2004, "The Ensembl automatic gene annotation system", *Genome Research*, Vol. 14, No. 5, pp. 942-950.

Datta K., J. L. Fuentes and J. R. Maddock, 2005, "The yeast GTPase Mtg2p is required for mitochondrial translation and partially suppresses an rRNA methyltransferase mutant, *mrm2*", *Molecular Biology of the Cell*, Vol. 16, No. 2, pp. 954-963.

De Hertogh B., E. Carvajal, E. Talla, B. Dujon, P. Baret, and A. Goffeau, 2002, "Phylogenetic classification of transporters and other membrane proteins from *Saccharomyces cerevisiae*", *Functional and Integrative Genomics*, Vol. 2, No. 4-5, pp. 154-70.

De Vit M. J., J. A. Waddle and M. Johnston, 1997, "Regulated nuclear translocation of the Mig1 glucose repressor", *Molecular Biology of the Cell*, Vol. 8, No. 8, pp. 1603-1618.

Delneri D., G. C. Tomlin, J. L. Wixon, A. Hutter, M. Sefton, E. J. Louis and S. G. Oliver, 2000, "Exploring redundancy in the yeast genome: an improved strategy for use of the cre-loxP system", *Gene*, Vol. 252, No. 1-2, pp. 127-135.

Delneri, D., 2004, "The Use of Yeast Mutant Collections in Genome Profiling and Large-Scale Functional Analysis", *Current Genomics*, Vol. 5, No. 1, pp. 59-65.

Deng M., Z. Tu, F. Sun and T. Chen, 2004, "Mapping Gene Ontology to proteins based on protein-protein interaction data", *Bioinformatics*, Vol. 20, No. 6, pp. 895-902.

Deng Y., T. He, Y. Wu, P. Vanka, G. Yang, Y. Huang, H. Yao and S. J. Brown, 2005, "Computationally analyzing the possible biological function of *YJL103C*--an ORF potentially involved in the regulation of energy process in yeast", *International Journal of Molecular Medicine*, Vol. 15, No. 1, pp. 123-127.

DeRisi J. L., V. R. Iyer and P. O. Brown, 1997, "Exploring the metabolic and genetic control of gene expression on a genomic scale", *Science*, Vol. 278, No. 5338, pp. 680-686.

DeRisi, J., B. Hazel, P. Marc, E. Balzi, P. Brown, C. Jacq and A. Goffeau, 2000, "Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants", *FEBS letters*, Vol. 460, pp. 156-160.

Dibrov E., S. Fu and B. D. Lemire, 1998, "The *Saccharomyces cerevisiae TCM62* Gene Encodes a Chaperone Necessary for the Assembly of the Mitochondrial Succinate Dehydrogenase (Complex II)", *Journal of Biological Chemistry*, Vol. 273, No.48, pp. 32042-32048.

Dolinski, K., R. Balakrishnan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. Nash, R. Oughtred, C. L. Theesfeld, G. Binkley, C. Lane, M. Schroeder, A. Sethuraman, S. Dong, S. Weng, S. Miyasato, R. Andrada, D. Botstein and J. M. Cherry, "Saccharomyces Genome Database" http://www.yeastgenome.org/ , latest time of access: December 2005.

Epstein, C. B., J. A. Waddle, W. H. IV, V. Davé, J. Thornton, T. L. Macatee, H. R. Garner and R. A. Butow, 2001, "Genome-wide Responses to Mitochondrial Dysfunction", *Molecular Biology of the Cell*, Vol. 12, pp. 297-308.

Erasmus D. J., G. K. van der Merwe and H. J. van Vuuren, 2003, "Genome-wide expression analyses: Metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress", *FEMS Yeast Research*, Vol. 3, No. 4, pp. 375-399.

EUROSCARF, 2003, http://www.uni-frankfurt.de/fb15/mikro/euroscarf/index.html, latest time of access: December 2005.

Feldmann, H., 2001, *Yeast Molecular Biology*, http://biochemie.web.med.uni-muenchen.de/Yeast_Biology/, latest time of access: December 2003.

Förster, J., I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen, 2003, "Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network", *Genome Research*, Vol. 13, pp. 244–253.

Galitski, T., A. J. Saldanha, C. A. Styles, E. S. Lander and G. R. Fink, 1999, "Ploidy regulation of gene expression", *Science*, Vol. 285, No. 5425, pp. 251-254.

Gancedo, J. M., 1998, "Yeast Carbon Catabolite Repression", *Microbiology and Molecular Biology Review*, Vol. 62, No. 2, pp. 334–361.

Gardner R. G, Z. W. Nelson and D. E. Gottschling, 2005, "Degradation-mediated protein quality control in the nucleus", *Cell*, Vol. 120, No. 6, pp. 803-15.

Gavin A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga, 2002, "Functional organization of the yeast proteome by systematic analysis of protein complexes", *Nature*, Vol. 415, No. 6868, pp. 141-147.

Geladi, P. and B. R. Kowalski, 1986, "Partial Least-Squares Regression: A Tutorial", *Analytica Chimica Acta*, Vol. 186, No. 1, pp. 1-17.

Giaever G., P Flaherty, J. Kumm, M. Proctor, C. Nislow, D. F. Jaramillo, A. M. Chu, M. I. Jordan, A. P. Arkin and R. W. Davis, 2004, "Chemogenomic profiling: identifying the functional interactions of small molecules in yeast", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 101, No. 3, pp. 793-798.

Gish, W., 1996-2004, http://blast.wustl.edu. Latest time of access: November, 2005.

Goffeau, A., B. G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnson, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin and S. G. Oliver, 1996, "Life with 6000 genes", *Science*, Vol. 274, No. 5287, pp. 546-567.

Gombert A. K, M. Moreira dos Santos, B. Christensen and J. Nielsen, 2001, "Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression", *Journal of Bacteriology*, Vol. 183, No. 4, pp. 1441-1451.

Griffin, T. J., S. P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood and R. Aebersold, 2002, "Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in *Saccharomyces cerevisiae*", *Molecular and Cellular Proteomics*, Vol. 1, No. 4, pp. 323-333.

Grotkjaer, T., 2005, *Bioinformatics Tools in Metabolic Engineering of Saccharomyces cerevisiae*, Ph.D. Thesis, Denmark Technical University.

Güldener U., M. Munsterkotter, G. Kastenmuller, N. Strack, J. van Helden, C. Lemer, J. Richelles, S. J. Wodak, J. Garcia-Martinez, J. E. Perez-Ortin, H. Michael, A. Kaps, E. Talla, B. Dujon, B. Andre, J. L. Souciet, J. De Montigny, E. Bon, C. Gaillardin and H. W. Mewes, "CYGD: the Comprehensive Yeast Genome Database", 2005, *Nucleic Acids Research*, Vol. 33, Database issue, pp. D364-368.

Güldener, U., S. Heck, T. Fiedler, J. Beinhauer and J. H. Hegemann, 1996, "A new efficient gene disruption cassette for repeated use in budding yeast", *Nucleic Acids Research*, Vol. 24, No. 13, pp. 2519–2524.

Gygi, S. P., Y. Rochon, B. R. Franza, and R. Aebersold, 1999, "Correlation between Protein and mRNA Abundance in Yeast", *Molecular and Cellular Biology*, Vol. 19, pp. 1720-1730.

Hallstrom, T. C. and W.S. Moye-Rowley, 2000, "Multiple Signals from Dysfunctional Mitochondria Activate the Pleiotropic Drug Resistance Pathway in *Saccharomyces cerevisiae*", *Journal of Biological Chemistry*, Vol. 275, No. 48, pp. 37347–37356,

Harrington, C. A., C. Rosenow and J. Retief, 2000, "Monitoring gene expression using DNA microarrays", *Current Opinion in Microbiology*, Vol. 3, No. 3, pp. 285-291.

Haurie, V., M. Perrot, T. Mini, P. Jeno, F. Sagliocco, H. Boucherie, 2001, "The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in *Saccharomyces cerevisiae*", *Journal of Biological Chemistry*, Vol. 276, No. 1, pp. 76-85.

Hayes, A., N. Zhang, J. Wu, P. R. Butler, N. C. Hauser, J. D. Hoheisel, F L. Lim, A. D. Sharrocks and S. G. Oliver, 2002, "Hybridization array technology coupled with chemostat culture: Tools to interrogate gene expression in *Saccharomyces cerevisiae*", *Methods*, Vol. 26, No. 3, pp. 281-290.

Hazbun, T. R., L. Malmstrom, S. Anderson, B. J. Graczyk, B. Fox, M. Riffle, B. A. Sundin, J. D. Aranda, W. H. McDonald, C. H.Chiu, B. E. Snydsman, P. Bradley, E. G. Muller, S. Fields, D. Baker, J. R. 3rd Yates and T. N. Davis, 2003, "Assigning function to yeast proteins by integration of technologies", *Molecular Cell*, Vol. 12, No. 6, pp. 1353-1365.

Hell K., J. Herrmann, E. Pratje, W. Neupert and R. A. Stuart, 1997, "Oxa1p mediates the export of the N- and C-termini of pCoxII from the mitochondrial matrix to the intermembrane space", *FEBS Letters*, Vol. 418, No. 3, pp. 367-370.

Ho Y., A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthiesen, R. C. Hendrickson,

F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. Hogue, D. Figeys and M. Tyers, 2002, "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry", *Nature*, Vol. 415, No. 6868, pp. 180-183.

Holstege, F. C., E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner, M. R. Green, T. R. Golub, E. S. Lander and R. A. Young, 1998, "Dissecting the regulatory circuitry of a eukaryotic genome", *Cell*, Vol. 95, No. 5, pp. 717-28.

Hopke, P. K., 2003, "The evolution of chemometrics", *Analytica Chimica Acta*, Vol. 500, pp. 365–377.

Hughes, T. R., M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard and S. H. Friend, 2000, "Functional Discovery via a Compendium of Expression Profiles", *Cell*, Vol. 102, No. 1, pp. 109-126.

Huh W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman and E. K. O'Shea, 2003, "Global analysis of protein localization in budding yeast", *Nature*, Vol. 425, No. 6959, pp. 686-691.

Hunte, C., J. Koepke, C. Lange, T. Rossmanith and H. Michel, 2000, "Structure at 2.3A resolution of the cytochrome bc_1 complex from the yeast Saccharomyces Cerevisiae co-crystallized with an antibody Fv fragment", *Structure*, Vol. 8, pp. 669-684.

Hutter, A. and S. G. Oliver, 1998, "Ethanol production using nuclear petite yeast mutants", *Applied Microbiology and Biotechnology*, Vol. 49, pp. 511-516.

Ideker, T., V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold and L. Hood, 2001, "Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network", *Science*, Vol. 292, No. 5518, pp. 929-933.

Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki, 2001, "A comprehensive two-hybrid analysis to explore the yeast protein interactome", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 98, No. 8, pp. 4569-4574.

Jiang, T, and A.E. Keating, 2005, "AVID: an integrative framework for discovering functional relationships among proteins", *BMC Bioinformatics*, Vol. 6, No. 1, p.136.

Johansson, D., P. Lindgren and A. Berglund, 2003, "A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription", *Bioinformatics*, Vol. 19, pp. 467-473.

Joshi, T., Y. Chen, J. M. Becker, N. Alexandrov and D. Xu, 2004, "Genome-scale gene function prediction using multiple sources of high-throughput data in yeast *Saccharomyces cerevisiae*", *OMICS*, Vol. 8, No. 4, pp. 322-333.

Kanehisa, M. and S. Goto, 2000, "KEGG: Kyoto Encyclopedia of ORFs and Genomes", *Nucleic Acids Research*, Vol. 28, pp. 27-30.

Kaniak, A., Z. Xue, D. Macool, J. H. Kim and M. Johnston, 2004, "Regulatory network connecting two glucose signal transduction pathways in *Saccharomyces cerevisiae*" *Eukaryotic Cell*, Vol. 3, No. 1, pp. 221-231.

Katz, M., T. Frejd, B. Hahn-Hagerdal and M. F. Gorwa-Grauslund, 2003, "Efficient anaerobic whole cell stereoselective bioreduction with recombinant *Saccharomyces cerevisiae*", *Biotechnology and Bioengineering*, Vol. 84, No. 5, pp. 573-82.

Katz, M., B. Hahn-Hägerdal and M. F. Gorwa-Grauslund, 2003, "Screening of two complementary collections of *Saccharomyces cerevisiae* to identify enzymes involved in stereo-selective reductions of specific carbonyl compounds: an alternative to protein purification", *Enzyme and Microbial Technology*, Vol. 33, No.2-3, pp. 163-172.

Khan, S., G. Situ, K. Decker and C. J. Schmidt, 2003, "GoFigure: automated Gene Ontology annotation", *Bioinformatics*, Vol. 19, No. 18, pp. 2484-2485.

King, O. D., J. C. Lee, A. M. Dudley, D. M. Janse, G. M. Church and F. P. Roth, 2003, "Predicting phenotype from patterns of annotation", *Bioinformatics*, Vol. 19, Supplementary issue 1, i183-i189.

Klein, C. J., L. Olsson and J. Nielsen, 1998, "Glucose control in *Saccharomyces cerevisiae*: the role of Mig1 in metabolic functions", *Microbiology*, Vol. 144, No.1, pp. 13-24.

Koski, L. B., M. W. Gray, B. F. Lang and G. Burger, 2005, "AutoFACT: an automatic functional annotation and classification tool", *BMC Bioinformatics*, Vol. 6, pp. 151.

Kourti, T. and J. F. MacGregor, 1995, "Process analysis, monitoring and diagnosis, using multivariate projection methods", *Chemometrics and Intelligent Laboratory Systems*, Vol. 28, pp. 3-21.

Kwasnicka D. A., A. Krakowiak, C. Thacker, C. Brenner and S. R. Vincent, 2003, "Coordinate expression of NADPH-dependent flavin reductase, Fre-1, and Hint-related 7meGMP-directed hydrolase, DCS-1", *Journal of Biological Chemistry*, Vol. 278, No. 40, pp. 39051-39058.

Lascaris, R., J. Piwowarski, H. van der Spek, J. T. de Mattos, L. Grivell and L. Blom, 2004, "Overexpression of *HAP4* in glucose-derepressed yeast cells reveals respiratory control of glucose-regulated genes", *Microbiology*, Vol. 150, pp. 929–934.

Lemaire, C., P. Hamel, J. Velours and G. Dujardin, 2000, "Absence of the mitochondrial AAA protease Yme1p restores F_0-ATPase subunit accumulation in an oxa1 deletion mutant of *Saccharomyces cerevisiae*". *Journal of Biological Chemistry*, Vol. 275, No. 31, pp. 23471-23475.

Lemaire, C., F. Guibet-Grandmougin, D. Angles, G. Dujardin and N. Bonnefoy, 2004, "A yeast mitochondrial membrane methyltransferase-like protein can compensate for oxa1 mutations", *Journal of Biological Chemistry*, Vol. 279, No. 46, pp. 47464-47672.

Letovsky, S. and S. Kasif, 2003, "Predicting protein function from protein/protein interaction data: a probabilistic approach", *Bioinformatics*, Vol. 19, Supplementary issue 1, pp. i197-i204.

Li, B. and B. Trueb, 2000, "DRG represents a family of two closely related GTP-binding proteins", *Biochimica et biophysica acta*, Vol. 1491, No. 1-3, pp. 196-204.

Lockhart, D. J. and E. A.Winzeler, 2000, "Genomics, gene expression and DNA arrays", *Nature*, Vol. 405, pp. 827-836.

Lodi, T. and B. Guiard, 1991, "Complex transcriptional regulation of the Saccharomyces cerevisiae *CYB2* gene encoding cytochrome b2: *CYP1*(*HAP1*) activator binds to the *CYB2* upstream activation site UAS1-B2", *Molecular and Cellular Biology*, Vol. 11, No. 7, pp. 3762-72.

Mandavilli, B.S., J. H. Santos and B. Van Houten, 2002, "Mitochondrial DNA repair and aging", *Mutation Research*, Vol. 509, No. 1-2, pp. 127-151.

Marks, V. D., G. K. van der Merwe and H. J. van Vuuren, 2003, "Transcriptional profiling of wine yeast in fermenting grape juice: regulatory effect of diammonium phosphate", *FEMS Yeast Research*, Vol. 3, No. 3, pp. 269-287.

Martin, D. E., and M. N. Hall, 2005, "The expanding TOR signaling network", *Current Opinion in Cell Biology*, Vol. 17, No. 2, pp. 158-166.

Martin, D. M., M. Berriman and G. J. Barton, 2004, "GOtcha: a new method for prediction of protein function assessed by the annotation of seven genomes", *BMC Bioinformatics*, Vol. 5, p. 178.

Martinez, M. J., S. Roy, A. B. Archuletta, P. D. Wentzell, S. S. Anna-Arriola, A. L. Rodriguez, A. D. Aragon, G. A. Quinones, C. Allen and M. Werner-Washburne, 2004, "Genomic analysis of stationary-phase and exit in *Saccharomyces cerevisiae*: gene expression and identification of novel essential genes", *Molecular Biology of the Cell*, Vol. 15, No. 12, pp. 5295-5305.

Mewes H. W., C. Amid, R. Arnold, D. Frishman, U. Güldener, G. Mannhaupt, M. Münsterkötter, P. Pagel, N. Strack, V. Stumpflen, J. Warfsmann and A. Ruepp, 2004, "MIPS: analysis and annotation of proteins from whole genomes", *Nucleic Acids Research*, Vol. 32, Database issue, pp. D41-44.

Mills R, M. Rozanov, A. Lomsadze, T. Tatusova and M. Borodovsky, 2003, "Improving gene annotation of complete viral genomes", *Nucleic Acids Research*, Vol. 31, No. 23, pp. 7041-7055.

Nabieva, E., K. Jim, A. Agarwal, B. Chazelle and M. Singh, 2005, "Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps", *Bioinformatics*, Vol. 21, Supplementary issue 1, i302-i310.

Navarro-Avino, J. P., R. Prasad, V. J. Miralles, R. M. Benito and R. Serrano, 1999, "A proposal for nomenclature of aldehyde dehydrogenases in Saccharomyces cerevisiae and characterization of the stress-inducible *ALD2* and *ALD3* genes", *Yeast*, Vol. 15, No. 10A, pp. 829-842.

Nielsen, J. and L. Olsson, 2002, "An expanded role for microbial physiology in metabolic engineering and functional genomics: moving towards systems biology", *FEMS Yeast Research*, Vol. 2, pp. 175-181.

Nobrega, F. G., M. P. Nobrega and A. Tzagoloff, 1992, "*BCS1*, a novel gene required for the expression of functional Rieske iron-sulfur protein in *Saccharomyces cerevisiae*", *EMBO Journal*, Vol. 11, No. 11, pp. 3821-3829.

Nyugen, D.V. and D.M. Rocke, 2002a, "Tumor classification by partial least squares using microarray gene expression data", *Bioinformatics*, Vol. 18, pp. 39-50.

Nyugen, D.V. and D.M. Rocke, 2002b, "Partial least squares proportional hazard regression for application to DNA microarray survival data", *Bioinformatics*, Vol. 18, pp. 1625-1632.

Ohlmeier, S., A. J. Kastaniotis, J. K. Hiltunen and U. Bergmann, 2004, "The yeast mitochondrial proteome, a study of fermentative and respiratory growth", *Journal of Biological Chemistry*, Vol. 279, No. 6, pp. 3956-3979.

Oliver, S. G., 1996, "From DNA sequence to biological function", *Nature*, Vol. 379, pp. 597-600.

Oliver, S. G., M. K. Winson, D. B. Kell and F. Baganz., 1998, "Systematic functional analysis of the yeast genome", *TIBTECH*, Vol. 16, pp. 373-378.

Oliver, S. G., 1999, "Redundancy reveals drugs in action", *Nature Genetics*, Vol. 21, pp. 245-246.

Onda, M., K. Ota, T. Chiba, Y. Sakaki and T. Ito, 2004, "Analysis of gene network regulating yeast multidrug resistance by artificial activation of transcription factors: involvement of Pdr3 in salt tolerance", *Gene*, Vol. 332, pp. 51-59.

Özcan, S. and M. Johnston, 1999, "Function and regulation of yeast hexose transporters", *Microbiology and Molecular Biology Reviews*, Vol. 63, No. 3, pp. 554-569.

Påhlman, A. K., K. Granath, R. Ansell, S. Hohmann and L. Adler, 2002, "The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress", *Journal Biological Chemistry*, Vol. 276, No. 5, pp. 3555-3563.

Palese, L. L., A. Gaballo, Z. Technikova-Dobrova, N. Labonia, A. Abbrescia, S. Scacco, L. Micelli and S. Papa, 2003, "Characterization of plasma membrane respiratory chain and ATPase in the actinomycete *Nonomuraea sp. ATCC 39727*", *FEBS Microbiology Letters*, Vol. 228, No. 2, pp. 233-239.

Palsson, B. and T. E. Allen, 2001, "Sequence-Based Analysis of Metabolic Demands for Protein Synthesis in Prokaryotes", *Journal of Theoretical Biology*, Vol. 213, pp. 73-88.

Pedruzzi, I, N. Burckert, P. Egger and C. De Virgilio, 2000, "*Saccharomyces cerevisiae* Ras/cAMP pathway controls post-diauxic shift element-dependent transcription through the zinc finger protein Gis1", *EMBO Journal*, Vol. 19, No. 11, pp. 2569-2579.

Peterson, L. E., 2003, "Partitioning large-sample microarray-based gene expression profiles using principal components analysis", *Computer Methods and Programs in Biomedicine*, Vol. 70, No. 2, pp. 107-119.

Piper, M. D., P. Daran-Lapujade, C. Bro, B. Regenberg, S. Knudsen, J. Nielsen and J. T. Pronk, 2002, "Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*", *The Journal of Biological Chemistry*, Vol. 277, No. 40, pp. 37001-37008.

Poyton, R. O., V. B. Burke, D. C. Raitt, L. A. Allen, and E. A. Kellogg, 1997, "Effects of oxygen concentration on the expression of cytochrome c and cytochrome c oxidase genes in yeast", *The Journal of Biological Chemistry*, Vol. 272, No. 23, pp. 14705-14711.

Preuss, M., K. Leonhard, K. Hell, R. A. Stuart, W. Neupert and J. M. Herrmann, 2001, "Mba1, a novel component of the mitochondrial protein export machinery of the yeast *Saccharomyces cerevisiae*", *Journal of Cell Biology*, 2001, Vol. 153, No. 5, pp. 1085-1096.

Przybyla-Zawislak, B., D. M. Gadde, K. Ducharme and M. T. McCammon, 1999, "Genetic and biochemical interactions involving tricarboxylic acid cycle (TCA) function using a collection of mutants defective in all TCA cycle genes", *Genetics*, Vol. 152, No. 1, pp. 153-66.

Raamsdonk, L.M., B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam and S. G. Oliver, 2001, "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations", *Nature Biotechnology*, Vol. 19, No. 1, pp. 45-50.

Raghevendran, V., A. K. Gombert, B. Christensen, P. Kotter and J. Nielsen, 2004, "Phenotypic characterization of glucose repression mutants of *Saccharomyces cerevisiae* using experiments with 13C-labelled glucose", *Yeast*, Vol. 21, No. 9, pp. 769-79.

Reekmans, R., K. De Smet, C. Chen, P. Van Hummelen and R. Contreras, 2005, "Old yellow enzyme interferes with Bax-induced NADPH loss and lipid peroxidation in yeast", *FEMS Yeast Research*, Vol. 5, No. 8, pp. 711-25.

Rep., M. and L. A. Grivell, 1996, "*MBA1* encodes a mitochondrial membrane-associated protein required for biogenesis of the respiratory chain", *FEBS Letters*, Vol. 388, No. 2-3, pp. 185-188.

Richard, G. F., C. Fairhead and B. Dujon, 1997, "Complete transcriptional map of yeast chromosome XI in different life conditions", *Journal of Molecular Biology*, Vol. 268, No. 2, pp. 303-321.

Ross-Macdonald, P., P. S. Coelho, T. Roemer, S. Agarwal, A. Kumar, R. Jansen, K. H. Cheung, A. Sheehan, D. Symoniatis and L. Umansky, 1999, "Large-scale analysis of the yeast genome by transpodensity oligonucleotide arrays", *Nature*, Vol. 402, pp. 413-418.

Rougemont, J and P. Hingamp, 2003, "DNA microarray data and contextual analysis of correlation graphs", *BMC Bioinformatics*, Vol. 4, No. 1, p. 15.

Rubinstein, R. and I. Simon, 2005, "MILANO--custom annotation of microarray results using automatic literature searches", *BMC Bioinformatics*. Vol. 6, No. 1, p. 12.

Schlitt, T., K. Palin, J. Rung, S. Dietmann, M. Lappe, E. Ukkonen and A. Brazma, 2003, "From gene networks to gene function", *Genome Research*, Vol. 12, pp. 2568-2576.

Shevchenko, A., O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie and M. Mann, 1996, "Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 93, No.25, pp. 14440-14445.

Sickmann, A., J. Reinders, Y. Wagner, C. Joppich, R. Zahedi, H. E. Meyer, B. Schonfisch, I. Perschil, A. Chacinska, B. Guiard, P. Rehling, N. Pfanner and C. Meisinger, 2003, "The proteome of *Saccharomyces cerevisiae* mitochondria", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 100, No. 23, pp. 13207-13212.

Steinmetz, L.M., C. Scharfe, A. M. Deutschbauer, D. Mokranjac, Z. S. Herman, T. Jones, A. M. Chu, G. Giaever, H. Prokisch, P. J. Oefner and R.W. Davis, 2002, "Systematic screen for human disease genes in yeast", *Nature Genetics*, Vol. 31, pp. 400-404.

Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster and E. D. Gilles, 2002, "Metabolic network structure determines key aspects of functionality and regulation", *Nature*, Vol. 420, pp. 190-193.

Stephanopoulos, G. N., A. A. Aristidou and J. Nielsen, 1998, *Metabolic Engineering*, Academic Press, San Diego.

Stuart R. A, C. M. Cruciat, K. Hell, H. Fölsch and W. Neupert, 1999, "BCS1p, an AAA-family member, is a chaperone for the assembly of the cytochrome bc_1 complex", *The EMBO Journal*, Vol. 18, No. 19, pp. 5226-5233.

Stuart R. A, C. M. Cruciat, S. Brunner, F. Baumann and W. Neupert, 2000, "The cytochrome bc_1 and cytochrome c oxidase complexes in yeast mitochondria", *The Journal of Biological Chemistry*, Vol. 275, No. 24, pp. 18093-18098.

Stuart, R., 2002, "Insertion of proteins into the inner membrane of mitochondria: the role of the Oxa1 complex", *Biochimica et Biophysica Acta (BBA) Molecular Cell Research*, Vol. 1592, No. 1, pp. 79-87.

Sumner, L. W., P. Mendes and R. A. Dixon, 2003, "Plant metabolomics: large-scale phytochemistry in the functional genomics era", *Phytochemistry*, Vol. 62, pp. 817–836.

Tadi, D., R. N. Hasan, F. Bussereau, E. Boy-Marcotte and M Jacquet, 1999, "Selection of genes repressed by cAMP that are induced by nutritional limitation in *Saccharomyces cerevisiae*", *Yeast*, Vol. 15, No. 16, pp. 1733-1745.

Tai, S. L., V. M. Boer, P. Daran-Lapujade, M. C. Walsh, J. H. de Winde, J. M. Daran and J. T. Pronk, 2005, "Two-dimensional transcriptome analysis in chemostat cultures. Combinatorial effects of oxygen availability and macronutrient limitation in *Saccharomyces cerevisiae*", *Journal of Biological Chemistry*, Vol. 280, No. 1, pp. 437-447.

Tamames, J., 2005, "Text Detective: a rule-based system for gene annotation in biomedical texts", *BMC Bioinformatics*, Vol. 6, Supplementary Issue 1, p. S10.

Tatusov, R. L., M. Y. Galperin, D. A. Natale and E. V. Koonin, 2000, "The COG database: a tool for genome-scale analysis of protein functions and evolution", *Nucleic Acids Research*, Vol. 28, No. 1, pp. 33-36.

Tran, H. G., D. J. Steger, V. R. Iyer and A.D. Johnson, 2000, "The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor", *EMBO Journal*, Vol. 19, No. 10, pp. 2323-2331.

Troyanskaya, O. G., K. Dolinski, A. B. Owen, R. B. Altman and D. Botstein, 2003, "A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*)", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 100, No. 14, pp. 8348-8353.

Trumpower B. L. and R. B. Gennis, 1994, "Energy transduction by cytochrome complexes in mitochondrial and bacterial respiration: The enzymology of coupling electron transfer reactions to transmembrane proton translocation", *Annual Reviews in Biochemistry*, Vol. 63, pp. 675-716.

Tzagoloff, A. 1995, "Ubiquinol-cytochrome-c oxidoreductase from *Saccharomyces cerevisiae*", *Methods in Enzymology*, Vol. 260, pp.51-63.

Tzagoloff, A., A. Barrientos, M. H. Barros, I. Valnot, A. Rötig, and P. Rustin, 2002, "Cytochrome oxidase in health and disease", *Gene*, Vol. 286, pp. 53-63.

Unno, K, P. R. Juvvadi, H. Nakajima, K. Shirahige and K. Kitamoto, 2005, "Identification and characterization of rns4/vps32 mutation in the RNase T1 expression-sensitive strain of *Saccharomyces cerevisiae*: Evidence for altered ambient response resulting in transportation of the secretory protein to vacuoles", *FEMS Yeast Research*, Vol. 5, No. 9, pp. 801-12.

Valencia, A., 2005, "Automatic annotation of protein function", *Current Opinion in Structural Biology*, Vol. 15, No. 3, pp. 267-274.

van den Berg, M. A., P. de Jong-Gubbels and H. Y. Steensma, 1998, "Transient mRNA responses in chemostat cultures as a method of defining putative regulatory elements: application to genes involved in *Saccharomyces cerevisiae* acetyl-coenzyme A metabolism", *Yeast*, Vol. 14, No. 12, pp. 1089-1104.

Vidan S. and M. Snyder, 2001, "Large-scale mutagenesis: yeast genetics in the genome era", *Current Opinion in Biotechnology*, Vol. 12, pp. 28–34.

Vinayagam, A, R. Konig, J. Moormann, F. Schubert, R. Eils, K. H. Glatting and S. Suhai, 2004, "Applying Support Vector Machines for Gene Ontology based gene function prediction", *BMC Bioinformatics*, Vol. 5, p. 116.

von Mering, C., L. J. Jensen, B. Snel, S. D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M. A. Huynen and P. Bork, 2005, "STRING: known and predicted protein-protein associations, integrated and transferred across organisms", *Nucleic Acids Research*, Vol. 33, Database issue, pp. D433-D437.

Walker, G. M., 1998, *Yeast Physiology and Biotechnology*, John Wiley and Sons, New York.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke and H. Bussey, 1999, "Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis", *Science,* Vol. 285, pp. 901-907.

Wold, S., M. Sjostrom and L. Eriksson, 2001, "PLS-regression: a basic tool of chemometrics", *Chemometrics and Intelligent Laboratory Systems*, Vol. 58, pp. 109–130

Wu, K., P. Wu and J. P. Aris, 2001, "Nucleolar protein Nop12p participates in synthesis of 25S rRNA in *Saccharomyces cerevisiae*", *Nucleic Acids Research*, Vol. 29, No. 14, pp. 2938-2949.

Wu, J., N. Zhang, A. Hayes, K. Panoutsopoulou and S. G. Oliver, 2004, "Global analysis of nutrient control of gene expression in *Saccharomyces cerevisiae* during growth and starvation", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 101, No. 9, pp. 3148–3153.

Xie, M. W., F. Jin, H. Hwang, S. Hwang, V. Anand, M. C. Duncan and J. Huang, 2005, "Insights into TOR function and rapamycin response: chemical genomic profiling by using a high-density cell array method", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 102, No. 20, pp. 7215-7120.

Yan Tong, A. H., M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, N. Pagé, M. Robinson, S. Raghibizadeh, C. W. V. Hogue, H. Bussey, B. Andrews, M. Tyers and C. Boone, 2001, "Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants", *Science*, Vol. 294, pp. 2364-2368.

Yeang, C. H., H. C. Mak, S. McCuine, C. Workman, T. Jaakkola and T. Ideker, 2000, "The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor", *EMBO Journal*, Vol. 19, No. 10, pp. 2323-2331.

Yeast Search for Transcriptional Regulators and Consensus Tracking (YEASTRACT), http://www.yeastract.com, latest time of access: December, 2005.

Young, E. T., K. M. Dombek, C. Tachibana and T. Ideker, 2003, "Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8", *Journal of Biological Chemistry*, Vol. 278, No. 28, pp. 26146-26158.

Zhou, X., M. C. Kao and W. H. Wong, 2002, "Transitive functional annotation by shortest-path analysis of gene expression data", *Proceedings of the National Academy of Sciences of the United States of the America*, Vol. 99, No. 20, pp. 12783-12788.

Zhou, Y., G. M. Huang and L. Wei, 2002, "UniBLAST: a system to filter, cluster, and display BLAST results and assign unique gene annotation", *Bioinformatics*, Vol. 18, No. 9, pp. 1268-1269.