

NETWORK TOPOLOGY AND DYNAMIC DATA ANALYSIS IN *SACCHAROMYCES*  
*CEREVISIAE*

by

Muhammed Erkan Karabekmez

B.S., Chemical Engineering, Boğaziçi University, 2004

M.S., Chemical Engineering, Boğaziçi University, 2010

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

2016

## ACKNOWLEDGEMENTS

In this chapter of my life a lot of things have changed around me, my perspective on the life and other people have evolved and while I am sitting on this milestone and looking back to sum up the last six years, all I can tell is that the life is a journey of seeking knowledge on which with each step the end of the tunnel seems to be getting farther. I will walk though...

I will be singing Prof. Betül Kırdar's praises who have guided me through this journey and without her this thesis could not be accomplished. I also need to thank to all instructors in my PhD-track and especially to Assoc. Prof. Kazım Yalçın Arga for his helpful discussions.

I would like to sincerely thank to KB441 folks; Duygu, Ayça, Ceyda, Elif, Esra, Serpil, Şebnem, Hilal, Saliha to whom I have been privileged to ask any question without any hesitation, they made the process much easier.

In the same years I have been in the ArGe, and I need to express my special thanks to the associates from the ArGe for the peaceful working environment.

I have spent hours of overtime at social associations voluntarily without which this thesis would have progressed faster, no regrets! I have to thank a lot of comrades who made the life more meaningful.

I have also a deep gratitude to all of my friends, number of whom is always complained by my wife. These guys have always supported me like brothers.

At the end of each day I end up at home. I believe that we are not created to be alone that is why the most important support I have had is always from my family who have become larger in these years; mom, dad, brothers, sister, four nieces, family in law, my dearest wife (or I should say my other half), Zeynep Hanne the cutest, and the expected son... There is no need for words.

Lastly I should acknowledge financial supports provided by TUBITAK through the project 106M444 and Bogazici University Research Fund (BAP) through the project 14A05D2.

## ABSTRACT

### NETWORK TOPOLOGY AND DYNAMIC DATA ANALYSIS IN *SACCHAROMYCES CEREVISIAE*

Biological systems which can be represented as networks and graphs are highly dynamic and responsive to environmental and genetic perturbations in a time dependent manner. These networks are hierarchically organized and consist of tightly clustered groups of proteins that work together as part of a biological process or a complex to achieve a specific function in a cell. With the emergence of high-throughput dynamic datasets, dynamic data analysis became a challenge in systems biology with the other challenges such as representation of biological systems as networks and elucidation of graph properties of these networks biologically and integration of multi -omics datasets in order to extract biologically meaningful results. The aim of this thesis is to develop a novel metric of centrality to identify biologically important nodes and to develop novel approaches to investigate dynamic datasets. In the first part, a novel global metric of centrality, weighted sum of loads eigenvector centrality (WSL-EC), counting all eigenvectors was proposed to identify essential and biologically central nodes. WSL-EC was found to outperform in capturing biologically central nodes, such as pathogen-interacting, HIV-1, cancer, ageing, and disease-related genes and genes involved in immune system process and related to autoimmune diseases in the human interactome compared with other metrics of centrality. In the second part dynamic transcriptional response of *S. cerevisiae* cells to doxorubicin, which is used as chemotherapeutic reagent in the treatment of different types of cancer, was monitored by quantification of RNA transcripts in cells which were grown in a chemostat fermenter, through microarray technology. Resulting dynamic transcriptome data were investigated by using different approaches and integrating interactome and regulome. The clustering and analysis of the transcriptomic response of *S. cerevisiae* cells to doxorubicin indicated that the genes involved in DNA replication, mismatched repair, cell cycle and base excision repair pathways were affected and several transcriptional factors were identified. In the third part the data collected from literature related to the transcriptional response of yeast cells to DNA damage was similarly investigated and compared with the response to doxorubicin.

## ÖZET

### AĞ İLİNGESİ VE *SACCHAROMYCES CEREVISIAE*'DE DEVİNGEN VERİ ANALİZİ

Ağ şeklinde de temsil edilebilen biyolojik sistemler oldukça devingen olup çevresel ve genetik değişikliklere karşı zamana bağlı tepki vermektedirler. Bu ağlar hiyerarşik bir örgütlenme içerisinde ve hücre içinde belirli işlevleri yerine getiren kompleks veya süreçlerin parçaları olarak çalışan birbirleriyle sıkı bir şekilde ilintili protein kümelerinden oluşurlar. Hızlı-tarama yöntemleri sonucu elde edilen devingen veri setlerinin ortaya çıkması ile birlikte devingen veri analizi sistem biyolojisi için önemli ve zorlu bir konu haline gelmiştir. Sistem biyolojisinin önündeki diğer önemli konular ise biyolojik sistemlerin ağ şeklinde temsil edilmesi ve ağ özelliklerinin biyolojik açıklamalarını ortaya çıkarmak ve çoklu –omik veri setlerinin birleştirilmesiyle biyolojik anlamlılığa sahip sonuçlara ulaşmak şeklindedir. Bu tezin amacı özgün bir merkezilik ölçüsü geliştirerek biyolojik öneme sahip noktaları belirlemek ve devingen veri setlerini incelemek için özgün yaklaşımlar geliştirmektir. İlk bölümde özgün bir merkezilik ölçüsü olarak bütün öz-yöneyle göz önüne alan, biyolojik açıdan merkezi özellikte olan ve hayati noktaları belirleyebilecek “ağırlıklı yükler toplamı öz-yöney merkezilik (AYTÖM)” geliştirilmiştir. AYTÖM’ün insan protein etkileşim ağına HIV-1, kanser, yaşlanma veya hastalıklarla ilgili genleri, bağışıklık sistemi sürecini etkileyen genleri ve özbağışıklık hastalıklarıyla ilgili genleri tespit etme konusunda diğer merkezilik ölçülerinden daha iyi bir performans gösterdiği görülmüştür. İkinci kısımda *S. cerevisiae* hücrelerinin farklı kanser tiplerinin tedavisinde kullanılan bir kemoterapi ilacı olan doxorubicin’e devingen gen anlatım tepkisi, sürekli fermentörde büyütülen maya hücrelerindeki RNA miktar tayininin mikrodizi teknoloji kullanılarak belirlenmesiyle ölçülmüştür. Elde edilen devingen gen anlatım verisi farklı yaklaşımlar kullanılarak, etkileşim ve düzenleyici -omik veri tümleştirmeleriyle incelenmiştir. *S. cerevisiae* hücrelerinin doxorubicin’e gen anlatımsal tepkisinin kümelenmesi ve incelemesi ile DNA replikasyonu, hatalı eşleşme onarımı, hücre döngüsü ve baz çıkartarak onarım yollarının etkilendiği ve bazı gen anlatım faktörleri belirlenmiştir. Üçüncü kısımda DNA hasarıyla ilgili literatürden toplanan veriler benzer şekilde analiz edilmiş ve doxorubicin’e verilen tepkiyle karşılaştırılmıştır.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
ABSTRACT .....	v
ÖZET .....	vi
LIST OF FIGURES .....	ix
LIST OF TABLES .....	xv
LIST OF SYMBOLS .....	xx
LIST OF ACRONYMS AND ABBREVIATIONS .....	xxi
1. INTRODUCTION .....	1
2. STUDIES ON PROTEIN-PROTEIN INTERACTION NETWORK TOPOLOGY TO IDENTIFY BIOLOGICALLY SIGNIFICANT PROTEINS .....	3
2.1. Background .....	3
2.2. Materials and Methods .....	6
2.2.1. Centrality Metrics and Network Parameters .....	6
2.2.2. Databases and the Novel Metric .....	9
2.3. Results .....	10
2.3.1. Description of the Novel Metric .....	11
2.3.2. Construction and Topological Analysis of the Networks .....	11
2.3.3. Network Integrity .....	16
2.3.4. Hubs .....	19
2.3.5. Biological Investigation of Hubs .....	23
2.3.6. Centrality and Lethality in Yeast .....	26
2.3.7. Biological and Topological Centrality in the Human Interactome .....	29
2.4. Discussion .....	32
2.5. Future Prospects .....	33
3. INVESTIGATION OF DYNAMIC RESPONSE OF YEAST CELLS TO A CHEMOTHERAPEUTIC AGENT: DOXORUBICIN .....	34
3.1. Background .....	34
3.2. Materials and Methods .....	36
3.2.1. Strain, Growth Conditions, Pulse Injection and Sampling .....	36

3.2.2. Sampling, RNA Isolation and Preprocessing the .....	37
3.2.3. Transcriptome Data Analysis .....	38
3.3. Results .....	40
3.3.1. Temporal organization of Global Transcriptional Response .....	41
3.3.2. Analysis of Global Transcriptional Response .....	42
3.3.3. Integrative approaches .....	51
3.3.4. Transcriptional Regulation .....	54
3.3.5. Differentially expressed genes .....	59
3.3.6. Development of an extended approach to identify differentially and significantly expressed genes .....	61
3.4. Discussion .....	87
3.4.1. Dynamic response of Yeast cells to doxorubicin pulse .....	87
3.4.2. Improvements provided by E-EDGE .....	91
3.4.3. Improvements provided by the reallocation pipeline in WGCNA .....	92
3.5. Future prospects .....	92
4. INVESTIGATION OF DYNAMIC RESPONSE OF YEAST CELLS TO DNA DAMAGE IN TRANSCRIPTOMIC LANDSCAPE .....	94
4.1. Background .....	94
4.2. Methods .....	97
4.2.1. Transcriptome Data .....	97
4.2.2. Transcriptome Data Analysis .....	97
4.3. Results .....	98
4.3.1. Identification of differentially and significantly expressed genes by E-EDGE .....	98
4.3.2. SOMs of differentially expressed genes identified by E-EDGE .....	99
4.3.3. WGCNA of differentially expressed genes identified by E-EDGE .....	105
4.3.4. Regulation of the response to DNA damage .....	117
4.3.5. NP analysis of the DNA damage .....	118
4.4. Discussion .....	123
4.5. Future prospects .....	126
5. CONCLUSION .....	128
REFERENCES .....	134

## LIST OF FIGURES

Figure 2.1.	Degree distribution (on left) and shortest path length distribution (on right) of Y1, Y2 and H networks .....	13
Figure 2.2.	Correlation profiles of protein-protein interactions and clustering coefficient variations with degree in (A) Y1, (B) Y2, and (C) H. Corresponding average clustering coefficient variations were fitted to power law line on a logarithmic scale (on the right) ....	14
Figure 2.3.	Average nearest neighbor degree as a function of degree. Distribution fits to power law with $R^2=0.934$ (slope of the line is $-0.597$ ) .....	16
Figure 2.4.	Change in efficiency and diameter by random and targeted node removals from (A) Y1, (B) Y2, and (C) H networks .....	17
Figure 2.5.	Red nodes are hubs defined by (A) DC, (B) EC, (C) BC and (D) WSL-EC in Y1 network .....	20
Figure 2.6.	Red nodes are hubs defined by (A) BC, (B) EC, (C) DC, (D) WSL-EC in the H network .....	21
Figure 2.7.	Distribution of the hubs (red nodes) defined by (A) DC, (B) EC, (C) BC, (D) WSL-EC in Y2 network .....	22
Figure 2.8.	Venn Diagram of Degree Central (DC), Eigenvector Central (EC) and WSL Eigenvector Central (WSL-EC) Hub Sets in (A) Y1, (B) Y2, and (C) H Networks .....	25

Figure 2.9.	Enrichment of essential genes within different hub sets of (A) Y1 and (B) Y2 networks .....	26
Figure 2.10.	Change in number of essential genes detected by centrality ranks and random ranks in (A) Y1 and (B) Y2 networks by jackknifing .....	27
Figure 2.11.	Change in number of biologically significant genes detected by centrality ranks and random ranks in the H network .....	30
Figure 3.1.	Hierarchical clustering of time points for time series transcriptome data of doxorubicin stress .....	41
Figure 3.2.	Variation of number of clusters and size of the largest cluster with minimum similarity in hierarchical clustering of microarray data of dynamic response to doxorubicin in yeast cells .....	42
Figure 3.3.	Hierarchical clustering dendrogram of dynamic response to doxorubicin in yeast cells. (A) is the largest cluster and (B) is the second largest cluster .....	43
Figure 3.4.	The variation of the average normalized expression profiles of the genes in A and B clusters (A and B respectively). The dashed lines represents standard deviation .....	43
Figure 3.5.	Splines of the expression profiles of the principal components identified by PCA. Percentages of the total variance explained by the principal components are written within the parenthesis .....	44
Figure 3.6.	SOMs of transcriptome data representing dynamic response to doxorubicin stress. Numbers beginning with 'c' are cluster IDs and the second numbers are the number of nodes in the clusters. Meta-clusters MC1 to MC6 were outlined by colored borders ....	47

Figure 3.7.	Fitness to scale-free nature (A) and mean connectivity (B) as a function of parameterization of the power function .....	48
Figure 3.8.	Hierarchical clustering dendrogram of 14 clusters by WGCNA ...	49
Figure 3.9.	Meta-analysis of the modules identified by WGCNA. Hierarchical organization of the modules (A), splines of eigengene profiles (B) and heat-map of the modules (C) .....	50
Figure 3.10.	Frequency distribution of PCCs and network properties of NP network with respect to PCC threshold .....	52
Figure 3.11.	Average expression profiles of the modules identified by NP analysis .....	54
Figure 3.12.	Dynamic regulatory map of yeast cells in response to doxorubicin pulse. X-axis was arranged as uniform sample points instead of real time scale in order to prevent visual ambiguity. Light-green dots indicate the location of bifurcations .....	55
Figure 3.13.	Bifurcation of green and violet branches 5 minutes after doxorubicin addition .....	56
Figure 3.14.	Bifurcation of the khaki branch 30 minutes after doxorubicin addition .....	57
Figure 3.15.	Bifurcation of the dark green branch 60 minutes after doxorubicin addition .....	57
Figure 3.16.	Expression profiles of Petrol Blue, Olive Green and Blue branches	58

Figure 3.17.	Venn diagram of significant genes identified by EDGE and used in DREM and SOMs analysis .....	61
Figure 3.18.	Frequency distribution of log <sub>2</sub> ratios of average expression levels of the 2856 genes resides in at least one of three sets identified in EDGE, SOMs and DREM analysis. The curve represents normal distribution and dashed line marks p-value = 0.05 .....	62
Figure 3.19.	Self-organizing maps of time series transcriptome data representing dynamic response of differentially expressed genes identified by E-EDGE .....	63
Figure 3.20.	Fitness to scale-free nature (A) and mean connectivity (B) as a function of parameterization of the power function .....	68
Figure 3.21.	Hierarchical clustering dendrogram of 12 clusters of E-EDGE genes by WGCNA .....	69
Figure 3.22.	Meta-analysis of the modules identified by WGCNA for E-EDGE genes. Hierarchical organization of the modules (A) and heat-map of the modules (B) .....	70
Figure 3.23.	Eigengene profiles of the modules identified by WGCNA of genes identified by E-EDGE (x-axis is sampling points and y-axis is fold change) .....	71
Figure 3.24.	The pipeline of reallocation of the genes in the modules identified by WGCNA for E-EDGE genes .....	72
Figure 3.25.	Average expression profiles of the positively correlated modules identified by WGCNA of the genes identified by E-EDGE (x-axis is sampling points and y-axis is fold change) .....	73

Figure 3.26.	Dynamic regulatory map of differentially expressed yeast genes identified by E-EDGE in response to doxorubicin pulse. X-axis was arranged as uniform sample points. Light-green dots indicate the location of bifurcations. Blue labels are the branch codes .....	79
Figure 3.27.	(A) Transcriptional profiles of A2100 (purple) and A2200 (green) branches and (B) two child branches of A2200; A2211 (brown) and A2212 (dark green) .....	81
Figure 3.28.	Frequency distribution of PCCs and network properties of NP network with respect to PCC threshold .....	83
Figure 3.29.	Average expression profiles of the modules identified by NP analysis of E-EDGE genes .....	84
Figure 4.1.	TF network of overlapping targets in response to MMS (Workman <i>et al.</i> , 2006) .....	95
Figure 4.2.	The procedure for dynamic transcriptional data analysis .....	96
Figure 4.3.	Self-organizing maps of expression profiles. The first six points and following four points represent the transcriptomic response to DNA damage introduced by MMS and UV respectively .....	98
Figure 4.4.	Fitness to scale-free nature (A) and mean connectivity (B) as a function of parameterization of the power function .....	105
Figure 4.5.	Hierarchical clustering dendrogram of genes identified by E-EDGE .....	105

Figure 4.6.	Meta-analysis of the modules identified by WGCNA for E-EDGE genes. Hierarchical organization of the modules (A) and heat-map of the modules (B) .....	106
Figure 4.7.	Average expression profiles of the positively correlated modules identified by WGCNA of the genes identified by E-EDGE (x-axis is sampling points and y-axis is fold change) .....	108
Figure 4.8.	Dynamic regulatory map of differentially expressed genes. X-axis was arranged as uniform sample points. The first six points and following four points represent the transcriptomic response to DNA damage introduced by MMS and UV respectively .....	117
Figure 4.9.	Average expression levels of R and B modules in; (A) 1-P and (B) 2-WT .....	118

## LIST OF TABLES

Table 2.1.	Topological characteristics of the H network .....	12
Table 2.2.	Percent of nodes removed to attain defined collapse .....	18
Table 2.3.	Network properties of Hub-Networks .....	22
Table 2.4.	R-AUC values for DC, BC, WSL-EC, EC and random curves for Y1 and Y2 networks .....	28
Table 2.5.	R-AUC values for DC, BC, WSL-EC, EC and random curves for the H network .....	31
Table 3.1.	Module ID's and sizes determined by WGCNA .....	48
Table 3.2.	Distribution of the nodes and anti-correlated edges in NP network	52
Table 3.3.	Number of nodes in 4 modules identified in NP .....	53
Table 3.4.	Selected GO biological process terms and pathways significantly associated with differentially expressed genes identified by EDGE	60
Table 3.5.	Selected GO biological process terms and pathways significantly associated with downregulated clusters; c0, c1 and c2. The number of the genes was given in parenthesis .....	64
Table 3.6.	Selected GO biological process terms and pathways significantly associated with the smallest clusters c3 and c4. The number of the genes was given in parenthesis .....	65

Table 3.7.	Selected GO biological process terms and pathways significantly associated with clusters c5 and c8 which shows an upregulation in long term. The number of the genes was given in parenthesis .....	65
Table 3.8.	Selected GO biological process terms and pathways significantly associated with the upregulated clusters c6 and c7. The number of the genes was given in parenthesis .....	66
Table 3.9.	Module ID's and sizes of E-EDGE genes determined by WGCNA	68
Table 3.10.	Module ID's and sizes of E-EDGE genes determined by WGCNA after <i>signedKME</i> pipeline .....	73
Table 3.11.	Selected GO biological process terms and pathways significantly associated with clusters M1*, M5* and M10* .....	74
Table 3.12.	Selected GO biological process terms and pathways significantly associated with clusters M2* and its negatively correlated counterpart M2*N .....	75
Table 3.13.	Selected GO biological process terms and pathways significantly associated with clusters M3* and M6* which have opposite expression profiles .....	76
Table 3.14.	Selected GO biological process terms and pathways significantly associated with clusters M4*, M7* and M11* .....	77
Table 3.15.	Selected GO biological process terms significantly associated with clusters M8*, M9* and M12* .....	78
Table 3.16.	Number of genes in each branch and transcriptional regulators of the branches identified by DREM .....	80

Table 3.17.	Distribution of the nodes and anti-correlated edges in NP network	83
Table 3.18.	Selected GO biological process terms significantly associated with KC and KI modules .....	85
Table 3.19.	Selected GO biological process terms and pathways significantly associated with TC and TI modules .....	86
Table 4.1.	Selected GO biological process terms and pathways significantly associated with up-regulated clusters; c0 and c1. The number of the genes was given in parenthesis .....	100
Table 4.2.	Selected GO biological process terms and pathways significantly associated with the clusters c3 and c4. The number of the genes was given in parenthesis .....	101
Table 4.3.	Selected GO biological process terms and pathways significantly associated with cluster c2. The number of the genes was given in parenthesis .....	102
Table 4.4.	Selected GO biological process terms and pathways significantly associated with the down-regulated cluster c5. The number of the genes was given in parenthesis .....	103
Table 4.5.	Selected GO biological process terms and pathways significantly associated with the down-regulated cluster c6, c7 and c8. The number of the genes was given in parenthesis .....	104
Table 4.6.	Module ID's and sizes of E-EDGE genes determined by WGCNA after <i>signedKME</i> pipeline .....	108

Table 4.7.	Selected GO biological process terms and pathways significantly associated with ME1 and ME6 modules .....	109
Table 4.8.	Selected GO biological process terms and pathways significantly associated with clusters ME2 and ME3 .....	111
Table 4.9.	Selected GO biological process terms and pathways significantly associated with ME4 and ME7 modules .....	112
Table 4.10.	Selected GO biological process terms and pathways significantly associated with ME5 and ME18 modules .....	113
Table 4.11.	Selected GO biological process terms and pathways significantly associated with ME8 and ME9 modules .....	114
Table 4.12.	Significantly enriched GO biological process terms of ME11 module .....	114
Table 4.13.	Significantly enriched GO biological process terms of ME12 module .....	115
Table 4.14.	Significantly enriched GO biological process terms of ME14 module .....	115
Table 4.15.	Selected GO biological process terms and pathways significantly associated with ME16 modules .....	116
Table 4.16.	Significantly enriched GO biological process terms of ME19 module .....	117
Table 4.17.	Topological properties of NP networks .....	119

Table 4.18.	Selected GO biological process terms and pathways significantly associated with R module in response to MMS .....	120
Table 4.19.	Selected GO biological process terms and pathways significantly associated with B module in response to MMS .....	121
Table 4.20.	Selected GO biological process terms and pathways significantly associated with R module in response to UV exposure .....	122
Table 4.21.	Selected GO biological process terms and pathways significantly associated with B module in response to UV exposure .....	123

## LIST OF SYMBOLS

$a_{ij}$	Elements of the adjacency matrix of interaction network
$BC_i$	Betweenness centrality of node $i$
$c_A$	number of genes in the branch $A$ which are regulated by $TF_x$
$C(k)$	Clustering coefficient of degree $k$
$c_S$	total number of genes regulated by $TF_x$ before the split
$DC_i$	Degree centrality of node $i$
$e_{ij}$	Interactions in the graph of a ppi
$EC_i$	Eigenvector centrality of node $i$
$E(G)$	Graph representing a ppi
$k$	Degree
$k(i)$	Degree of node $i$
$l_i$	Number of links among the neighbors of the node $i$
$n_A$	number of genes in the branch $A$
$q_i$	Probability to have a node with degree $i$
$r$	Assortativity coefficient
$SC_i$	Subgraph centrality of node $i$
$spl_{ij}$	Shortest path length node $j$ and $i$
$TF_x$	A specific TF
$\sigma_{DC}$	Standard deviation of degree centralities
$\sigma_{jk}$	Total number of shortest paths between node $j$ and $k$
$\sigma_{jk}(i)$	Number of shortest paths between node $j$ and $k$ that passes through the node $i$
$\sigma_q$	Standard deviation of degree centralities
$\varepsilon$	Efficiency
$\lambda$	Eigenvalue
$\rho$	Network density
$v_j$	$j^{th}$ eigenvector

## LIST OF ACRONYMS AND ABBREVIATIONS

AC/MS	Affinity capture mass spectrometry
AUC	Area under curve
BC	Betweenness centrality
BC-HN	Betweenness centrality hub network
BioGRID	Biological general repository for interaction datasets
CC	Clustering coefficient
COSMIC	Catalogue of somatic mutations in cancer
cpl	Characteristic path length
D	Diameter
DAVID	Database for annotation, visualization and Integrated discovery
DC	Degree centrality
DC-HN	Degree centrality hub network
DEG	Database of essential genes
DIP	Database of interacting proteins
DNA	Deoxyribonucleic acid
DREM	Dynamic regulatory event miner
EC	Eigenvector centrality
EC-HN	Eigenvector centrality hub network
EDGE	Extraction of differential gene expression
E-EDGE	Extended extraction of differential gene expression
GenAge	Ageing gene database
GO	Gene ontology
HCE	Hierarchical clustering explorer
HIV-1	Human immunodeficiency virus 1
HPRD	Human protein reference database
I-AUC	Ideal area under curve
KEGG	Kyoto encyclopedia of genes and genomes
L	Liter
MINT	Molecular interaction database

MIPS	Mammalian protein-protein interaction database
MMS	Methyl methanesulfonate
N	Number of genes
NP	Negative/positive
PCA	Principal component analysis
PCC	Pearson correlation coefficient
ppi	Protein-protein interaction
R-AUC	Ratio of area under curve
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SC	Subgraph centrality
SOMs	Self-organising maps
spl	Shortest path length
STRING	Search tool for the retrieval of interacting genes/proteins
TAP/MS	Tandem affinity purification coupled to mass spectrometry
TOM	Topological overlap matrix
TF	Transcription factor
UV	Ultra violet
WGCNA	Weighted gene co-expression network analysis
WSL-EC	Weighted sum of loads eigenvector centrality
WSL-EC-HN	Weighted sum of loads eigenvector centrality hub network
Y2H	Yeast two-hybrid
YPD	Yeast extract-peptone-dextrose

## 1. INTRODUCTION

High-throughput technologies rapidly changed scientific perspectives in many fields leading the scientific community to a new era of data intensive age. This milestone was considered as important as invention of printing press (Tansley *et al.*, 2009). Evaluating, modeling, summarizing mass amount of data and extracting information are the challenges in this new era. This looking-for-a-needle-in-a-haystack mission requires gathering of diverse scientists from different fields like computing, mathematics, statistics and engineering in collaboration.

Implications of this progress in life sciences came with systems biology as a new branch of biology. Systems biology is mainly an attempt to combine diverse datasets and analyze them at once in order to understand how pieces work in an integrated system. Mathematical approaches from graph theory, machine learning and statistics have been implemented on biological datasets within this context.

Medical sciences, pharmacology, agricultural sciences, industrial biotechnology and all other life sciences resort to systems biology to develop targeted drugs, personalized medicines, advanced therapies, cell factories and yield in agriculture (Mustacchi *et al.*, 2006).

The genomic complexity increases with the complexity of the organisms however the difference was found to be less than expected. After publication of human genome sequence in 2001 the estimated number of genes in human was dropped drastically (Perteau *et al.*, 2010). It is now known that there are many conserved genes, proteins and pathways between human and even single-cell organisms. That is why yeast is a well studied model organism which has been used as a template in almost all life sciences for projections to other organisms.

Genome scale simultaneous high-throughput measurements of transcripts, metabolites, proteins, interactions and sequences provide omics datasets. Metabolomics can be used to infer fluxomics and outputs of the cell. Transcriptomics can be helpful in order to

unveil dynamics of transcriptional responses. Interactomics can be used to infer signaling pathways, central players or modular hierarchy. Integration of combinations of these omics datasets provide wide range of context-specific approaches.

In this thesis, genome scale datasets were analyzed in order to unveil transcriptional dynamics of yeast cells in response to genotoxic agents. Methodologies and metrics related to network topology and dynamic data analysis were tested, compared and novel improvements and novel metrics were suggested.

In the first part of the thesis (chapter 2), a novel metric of centrality—weighted sum of loads eigenvector centrality (WSL-EC)—based on graph spectra was defined and its performance in identifying topologically and biologically important nodes was comparatively investigated with common metrics of centrality in three real biological networks.

In the following part of the thesis (chapter 3), experiments were conducted where yeast cells were subjected to a chemical stress by introducing a chemotherapeutic drug; doxorubicin. Transcriptional re-arrangements that take place after the introduction were analyzed by well-established methods and some modified and developed versions of these methods. Integrative approaches were also utilized in order to assess depth in the response regarding physical interactions and regulatory associations. A commonly used existing approach for the identification of differentially expressed gene sets from dynamic data sets was modified to include a larger significant data set.

In the last part of the study (chapter 4), time series microarray datasets collected after a genotoxic stress in *Saccaromyces cerevisiae* were selected from the literature. Differentially expressed genes were identified by E-EDGE approach which was defined in the previous section. Differentially expressed genes within the selected datasets were analyzed by the pipeline constructed in the second chapter. Deciphered transcriptional response to DNA damage was comparatively investigated with the transcriptional response to doxorubicin.

## **2. STUDIES ON PROTEIN-PROTEIN INTERACTION NETWORK TOPOLOGY TO IDENTIFY BIOLOGICALLY SIGNIFICANT PROTEINS**

Topological centrality in protein interaction networks and its biological implications have widely been investigated in the past. In the present study, a novel metric of centrality—weighted sum of loads eigenvector centrality (WSL-EC)—based on graph spectra is defined and its performance in identifying topologically and biologically important nodes is investigated in comparison with common metrics of centrality in three real networks.

### **2.1. Background**

The understanding of life at the molecular level has impressively increased in the last half-century owing to technological advances such as microarrays, mass spectrometry and next generation sequencing. In addition to these technologies, which helped in the identification and quantification of biological molecules at the whole-genome level, high-throughput technologies were also developed to measure physical protein–protein, protein–DNA or RNA, and enzyme–metabolite interactions. All these interactions can be represented as networks or graphs, which can provide a good basis for modelling molecular interactions, integrating several sets of omics data and interpreting the overall physical and functional landscape of cellular function. These developments have provided the scientific basis of a new field known as network biology, which combines systems biology, graph theory, and statistical and computational analysis (Barabási and Oltvai, 2004).

Protein–protein interaction (ppi) networks at the whole genome level (interactome) are considered an important source to be investigated to obtain further information about cellular function. These networks are hierarchically organized and consist of tightly clustered groups of proteins that work together as part of a biological process or a complex to achieve a specific function in a cell. Protein–protein interactions were identified using yeast two-hybrid (Y2H) screening, tandem affinity purification coupled to mass spectrometry (TAP/MS) and affinity capture mass spectrometry (AC/MS) in several model organisms,

including humans, and are deposited in publicly available databases such as the Database of Interacting Proteins (DIP) (Xenarios *et al.*, 2000), Human Protein Reference Database (HPRD) (Peri *et al.*, 2003), Biological General Repository for Interaction Datasets (BioGRID) (Stark *et al.*, 2006) and STRING database (Jensen *et al.*, 2009).

Graph-theoretical analysis of these networks has revealed a strong correlation between the topological characteristics of cellular networks and cellular function. Early studies indicated the scale-free topology of protein–protein interaction networks, which consisted of a small number of hubs with many interactions. Today, molecular interaction networks are considered not to be scale-free but are generally heavy-tailed, consisting of few hubs and many low-degree nodes (Roy, 2012; De Lomana *et al.*, 2010).

It has been reported that some nodes are more important or *central* than the others in protein-protein interaction networks and the survival of an organism depends more on the few highly connected central nodes (Jeong *et al.*, 2001). This centrality-lethality relationship was later confirmed by several other studies in yeast and in other organisms (Batada *et al.*, 2006; Yu *et al.*, 2007; Hahn and Kern, 2005).

Further studies showed that the essentiality of proteins was linked to their involvement in the essential functional modules rather than their centrality (Zotenko *et al.*, 2008). The essentiality of the high degree nodes was explained by the fact that these proteins are more likely to interact with essential complexes and the removal of these nodes lead to the disruption of these complexes (Wang *et al.*, 2009; Ryan *et al.*, 2013).

Furthermore, the topological analysis of protein–protein interaction networks has provided a deeper understanding of biological systems, leading to the functional annotation of unknown genes or identification of drug targets or disease-related proteins and pathways (Wang *et al.*, 2011; Milenković *et al.* 2011). It has been reported that biologically important proteins in aging, cardiovascular disorders, metabolic disorders, cancer and infectious diseases have some topological centrality in the human interactome (Ideker and Sharan, 2008; Jonsson and Bates, 2006; Yildirim *et al.*, 2007; Ferrarini *et al.*, 2005; Dyer *et al.*, 2008).

A number of different metrics of topological centrality have been described to define the centrality of nodes, such as degree centrality (DC), which is the number of edges; betweenness centrality (BC), which is the fraction of shortest paths that pass through a node; and eigenvector centrality (EC). Although DC is the most commonly used measure of centrality, it can only give information about the local topology of a node. BC was used to determine bottleneck nodes that are low-degree but detrimental for the organism when removed (Yu *et al.*, 2007). EC, as a global metric of centrality, can explain latent topology by not only local connectivity but also the connectivity of the neighboring nodes (Wang *et al.*, 2011; McDermott *et al.*, 2012). Although EC is not a local metric of centrality like DC, it is limited to the first principal of graph spectra and therefore is not a descriptive metric for the peripheral modules in a network (Aguirre *et al.*, 2013).

Other measures of centrality such as subgraph centrality (Estrada and Rodriguez-Valazquez, 2005) (SC), which accounts for all graph spectra instead of only the first principal, have also been proposed to represent the number of short walks that start and end at the node of interest. The drawback of SC is that it converges to EC when the largest eigenvalue breaks away from the second (Benzi and Klymko, 2015).

Other measures of centrality, such as coreness centrality (Wuchty and Almaas, 2005), bipartivity (Estrada, 2006), graphlet degree centrality (Milenković *et al.*, 2011), node hierarchy (Bhardwaj *et al.*, 2010) and linear combination of different metrics (Roy, 2012), have also been proposed to improve the predictability of cellular functions or biologically central nodes in health or disease.

Different measures of centrality have been extensively used and compared for the topological analysis of biological networks (Filkov *et al.*, 2009; Roy and Filkov, 2009). It has been observed that different metrics of centrality can be important in different instances. Therefore, the development and application of different metrics are considered to be important in the topological analysis and modelling of networks in systems biology, in order to improve the predictability of cellular functions or biologically central nodes in health or disease.

## 2.2. Materials and Methods

### 2.2.1. Centrality Metrics and Network Parameters

For any graph,  $G$  there is a corresponding adjacency matrix,  $A$  (Equation 2.1).

$$A = (a_{ij})$$

$$(a_{ij}) = \begin{cases} 1 & e_{ij} \in E(G) \\ 0 & e_{ij} \notin E(G) \end{cases} \quad (2.1)$$

Degree centrality ( $DC_i$ ) is the number of direct neighbors (Equation 2.2) (Dong and Horvath, 2007).

$$DC_i = \sum_{j=1}^N a_{ij} \quad (2.2)$$

where  $N$  is the number of nodes in the network.

Betweenness centrality ( $BC_i$ ) of a node  $i$  is the ratio of shortest paths between any node couples that passes through node  $i$  (Equation 2.3) (Freeman, 1977).

$$BC_i = \sum_{j \neq k \neq i} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \quad (2.3)$$

where  $\sigma_{jk}(i)$  is the number of shortest paths between node  $j$  and  $k$  that passes through the node  $i$  and  $\sigma_{jk}$  is the total number of shortest paths between node  $j$  and  $k$ .

Eigenvector centrality ( $EC_i$ ) of a node  $i$  is the  $i^{th}$  component of the principal eigenvector of the adjacency matrix  $A$ . Non-zero vectors which satisfies Equation 2.4 are called eigenvectors where  $\lambda$  values are scalar and called eigenvalues.

$$Av = \lambda v \quad (2.4)$$

Then  $EC_i$  can be calculated as in Equation 2.5.

$$EC_i = v_1(i) \quad (2.5)$$

where  $v_1$  is the first principal component, *i.e.* the eigenvector corresponding to the largest eigenvalue (Bonacich, 1987).

Subgraph centrality ( $SC_i$ ) of a node  $i$  is the weighted sum of closed walks starting and ending at the node  $i$ , where short walks have higher weights with respect to longer walks (Estrada and Rodriguez-Valazquez, 2005) (Equation 2.6).

$$SC_i = \sum_{j=1}^N [v_j(i)]^2 e^{\lambda_j} \quad (2.6)$$

Network density can be defined as the ratio of number of links in a graph to number of maximum possible links (Equation 2.7) (Dong and Horvath, 2007).

$$\text{Network density} = 2 \frac{\sum_{i=1}^N \sum_{j=1}^N a_{ij}}{N(N-1)} \quad (2.7)$$

Network centralization is a measure of network compactness and it can be defined based on different centrality measures (Freeman 1978). In this work it was calculated by using Cytoscape which bases on degree centrality (Equation 2.8) (Dong and Horvath, 2007).

$$\text{Network centralization} \cong \frac{\max(DC) - \text{mean}(DC)}{N} \quad (2.8)$$

Coefficient of variation of degree distribution is defined as network heterogeneity (Equation 2.9) (Dong and Horvath, 2007).

$$\text{Network heterogeneity} = \frac{\sigma_{DC}}{\text{mean}(DC)} \quad (2.9)$$

Characteristic path length ( $cpl$ ) is the average of the shortest path lengths ( $spl$ ) between all possible node couples (Equation 2.10) (Assenov *et al.*, 2008).

$$cpl = 2 \frac{\sum_{i=1}^N \sum_{j=1}^N spl_{ij}}{N(N-1)} \quad (2.10)$$

Diameter is the longest shortest path between any pair of nodes. It gives the size of the largest connected part of the network (Equation 2.11) (Assenov *et al.*, 2008).

$$D = \max(spl_{ij}) \quad (2.11)$$

Clustering coefficient of a node is the ratio of links among its neighbors to the maximum possible number of links among its neighbors. Clustering coefficient of a network is the average clustering coefficient of all nodes (Equation 2.12) (Dong and Horvath, 2007).

$$CC = \frac{1}{N} \sum_{i=1}^N \frac{2l_i}{DC_i(DC_i - 1)} \quad (2.12)$$

where  $l_i$  is the number of links among the neighbors of the node  $i$ .

Assortativity coefficient is the slope of the line fitted to degree correlation distribution (Equation 2.13) (Newman, 2002).

$$r = \frac{1}{\sigma_q^2} \sum_{ij} ij(e_{ij} - q_i q_j) \quad (2.13)$$

where  $e_{ij}$  is the probability of having a link between a node with degree  $i$  to another node with degree  $j$ ,  $q_i$  is the probability to have a node with degree  $i$  and  $\sigma_q^2$  is the variance of the degree distribution.

Efficiency is the sum of reciprocal of  $spl$  between all possible node couples in a network (Equation 2.14) (Crucittia *et al.*, 2004).

$$\varepsilon = \sum_{i=1}^N \sum_{j=1}^N \frac{1}{spl_{ij}} \quad (2.14)$$

In this work centralities and parameters were calculated by either network analyzer plug-in of Cytoscape (Assenov *et al.*, 2008) or MATLAB R2012a software.

### 2.2.2. Databases and the Novel Metric

Global physical protein-protein interaction network of yeast (Y1) consisting of 5,891 nodes and 76,463 bidirectional edges without self-loops was downloaded from the BioGRID database (version 3.1.94) (Stark *et al.*, 2006).

The human interactome, H, was downloaded from the BioGRID database (version 3.2.108) (Stark *et al.*, 2006). H is composed of 15,192 nodes and 126,572 physical binary interactions without self-loops.

The functional ppi network, Y2, which consists of 1,792 proteins related to glucose processes and 6,919 bidirectional edges, was constructed (Yuzuak, 2012) from the STRING database (Jensen *et al.*, 2009) v8.3 with a confidence score of  $\geq 0.999$  using a selective permeability algorithm (Arga *et al.*, 2007), starting with 108 proteins that are associated with glucose metabolic process.

The novel spectral measure of centrality (WSL-EC) was defined as the weighted sum of loads of all principals of the graph spectra. Corresponding eigenvalues were used as weights and absolute values of all loads and weights were considered (Equation 2.15).

$$WSL-EC_i = \sum_{j=1}^N |\lambda_j| \cdot |v_{ij}| \quad (2.15)$$

where  $N$  is the number of nodes in the network,  $\lambda_j$  is the  $j^{th}$  eigenvalue of the adjacency matrix and  $v_{ij}$  is the load of the  $i^{th}$  node to the  $j^{th}$  principal of the graph spectra.

A total of 2417 HIV-1-interacting proteins (Fu *et al.*, 2009), 292 ageing-related genes from the Ageing Gene Database (GenAge) (Tacutu *et al.*, 2013), 407 cancer-related genes collected from the Catalogue of Somatic Mutations in Cancer (COSMIC) (Forbes *et al.*, 2008), 506 pathogen-interacting proteins (Dyer *et al.*, 2008), 1485 disease-related genes (Goh *et al.*, 2007), 1811 genes involved in immune system process (Ashburner *et al.*, 2000) and 143 genes related to autoimmune diseases (Cariaso and Lennon, 2012) were used as biologically important proteins.

Significantly associated Gene Ontology biological process terms ( $p$ -value  $< 0.001$ ) were determined by Gorilla (Boyle *et al.*, 2004).

Pathways that are related to the hubs were identified by KEGG release 76.0 (Kanehisa and Goto, 2000).

## 2.3. Results

The new measure of centrality, weighted sum of loads eigenvector centrality (WSL-EC), counts all eigenvectors using a different, simpler weighting strategy in order to capture topologically important nodes not only from the densely populated but also from the less densely populated and peripheral parts of the human network. The performance of WSL-EC in the identification of topologically important nodes that contribute to the integrity of network and to capture essential or biologically central nodes were tested in three real networks (*i*) yeast global protein-protein interaction network (*ii*) a functional subnetwork of yeast protein-protein interaction network and (*iii*) human global protein-protein interaction network. The performance of this newly introduced measure of centrality was compared with that of degree centrality (DC), betweenness centrality (BC), eigenvector centrality (EC) and subgraph centrality (SC).

### 2.3.1. Description of the Novel Metric

A novel metric of centrality called weighted sum of loads eigenvector centrality (WSL-EC) was developed. WSL-EC is designed as the weighted sum of loads of each node to each eigenvector. The absolute values of the loads were used as signs to indicate only the

direction, not the significance, of the load. Instead of exponentials of eigenvalues,  $e^{\lambda_i}$ , eigenvalues themselves,  $\lambda_i$ , were used as weights to prevent dominance of the first principal. Then its efficiency and power to identify central proteins were analyzed and compared with other commonly and extensively used metrics such as degree centrality (DC), betweenness centrality (BC), subgraph centrality (SC) and eigenvector centrality (EC).

### 2.3.2. Construction and Topological Analysis of the Networks

Physical protein-protein interaction network of yeast (Y1) consisting of 5,891 nodes and 76,463 bidirectional edges, glucose metabolism related functional protein-protein interaction network of yeast (Y2) consisting of 1,792 nodes and 6,919 bidirectional interactions and human global protein-protein interaction network (H) consisting of 15,192 nodes and 126,572 binary interactions without self-loops were constructed as described in Materials and Methods.

The power law distribution, hierarchical modularity and degree correlations on connected nodes are parameters that are commonly used to characterize the organization of a biological network (Hao *et al.*, 2012). Topological characteristics of three biological protein-protein interaction networks (Y1, Y2 and H) were calculated to determine the organization of the network (Table 2.1). All networks were found to have a degree distribution that fit to power law ( $R^2 \geq 0.844$ ) indicating scale-free nature of the networks (Figure 2.1).

Analysis and comparison of topological characteristics of the networks constructed in this study indicated that Y1 is the most compact and Y2 is the least compact networks. Y2 is probably the most modular network when compared to Y1 and H networks. The largest diameter were identified in Y2 which is consisting of lower number of nodes. This may indicate that it is less robust than Y1 and H networks. This analysis of network parameters, degree distributions and *spl* distributions suggest that H and Y1 networks have similar network architecture which is slightly distinct than that of the Y2 network (Table 2.1).

Table 2.1. Topological characteristics of the H network.

Network	Y1	Y2	H
Number of Nodes	5,891	1,792	15192
Number of Edges	76,463	6,919	126,572
Density ( $\rho$ )	0.004	0.004	0.001
Diameter (D)	6	21	8
Average Clustering Coefficient (CC)	0.251	0.494	0.256
Characteristic Path Length ( <i>cpl</i> )	2.645	7.388	2.670
Network Centralization	0.433	0.029	0.630
Network Heterogeneity	2.421	1.074	5.488

The topological characteristics of the H network indicate that although the nodes are loosely connected ( $\rho = 0.001$ ) and the diameter is higher than that of random networks of the same size and density (p-value  $< 0.01$ ), the average shortest path length (*cpl*) is lower and both network centralization and network heterogeneity are higher than these characteristics in a randomly wired network with the same number of nodes and edges (all p-values are less than 0.01). Together with findings that revealed the shortest path length (*spl*) distribution, more than 99% of the *spl*'s are 2, 3 or 4 steps long (Figure 2.1), these numbers indicate the presence of super-hubs, which are nodes with off-the-scale high connectivity. In addition, the non-randomly high clustering coefficient (p-value  $< 0.01$ ) that was observed is a sign of the modular organization of the network (Table 2.1).

The variation of the clustering coefficient is considered to be one of the most commonly used parameters for identifying the hierarchical modular structure of biological networks and it has been shown that the existence of super-hubs and degree correlation also affects the variation of the clustering coefficient with degree (Pastor-Satorras et al., 2001).

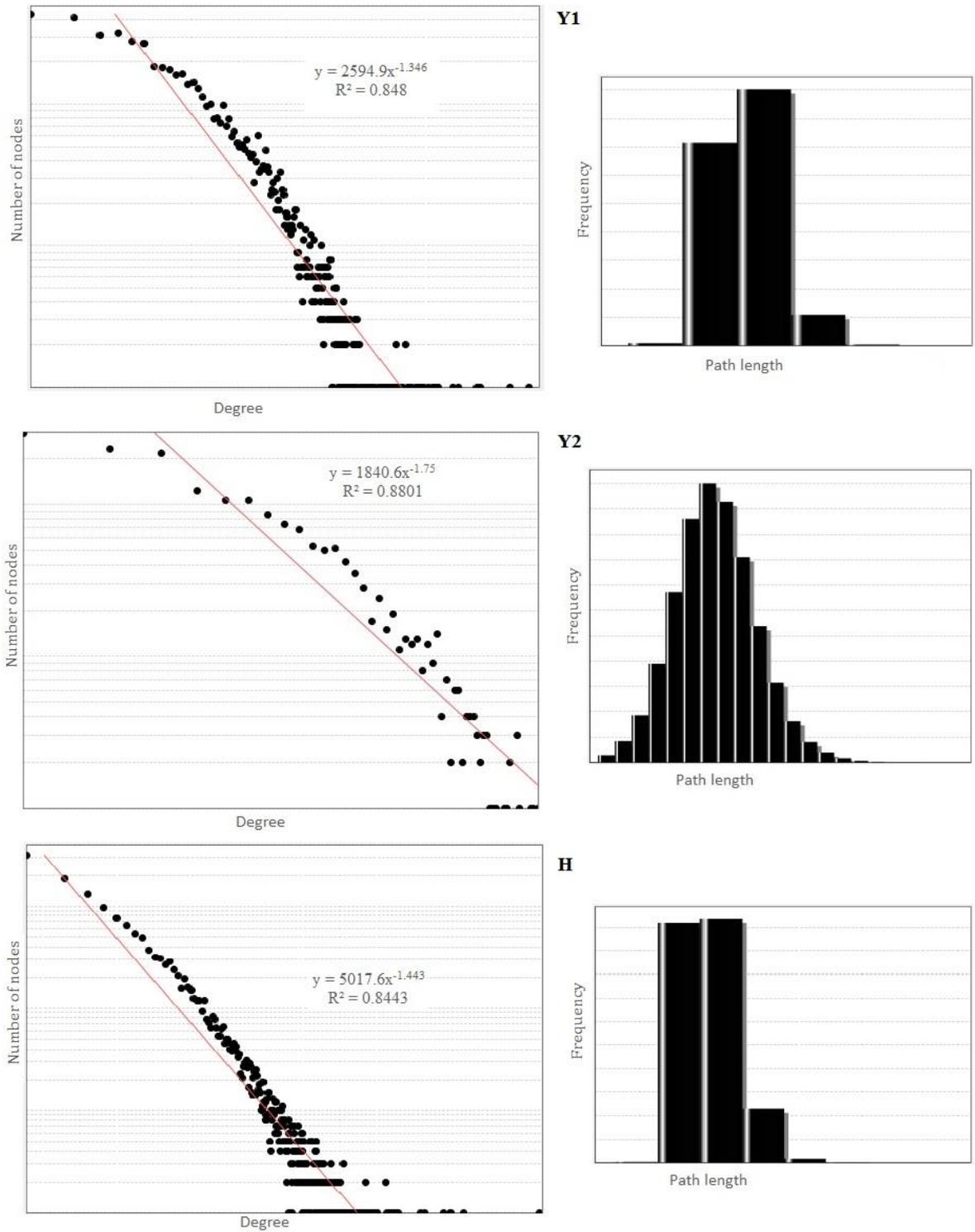


Figure 2.1. Degree distribution (on left) and shortest path length distribution (on right) of Y1, Y2 and H networks.

In order to reveal the hierarchical architectural structure of the networks, metrics of assortativity and variation of the clustering coefficient with degree were investigated (Figure 2.2).

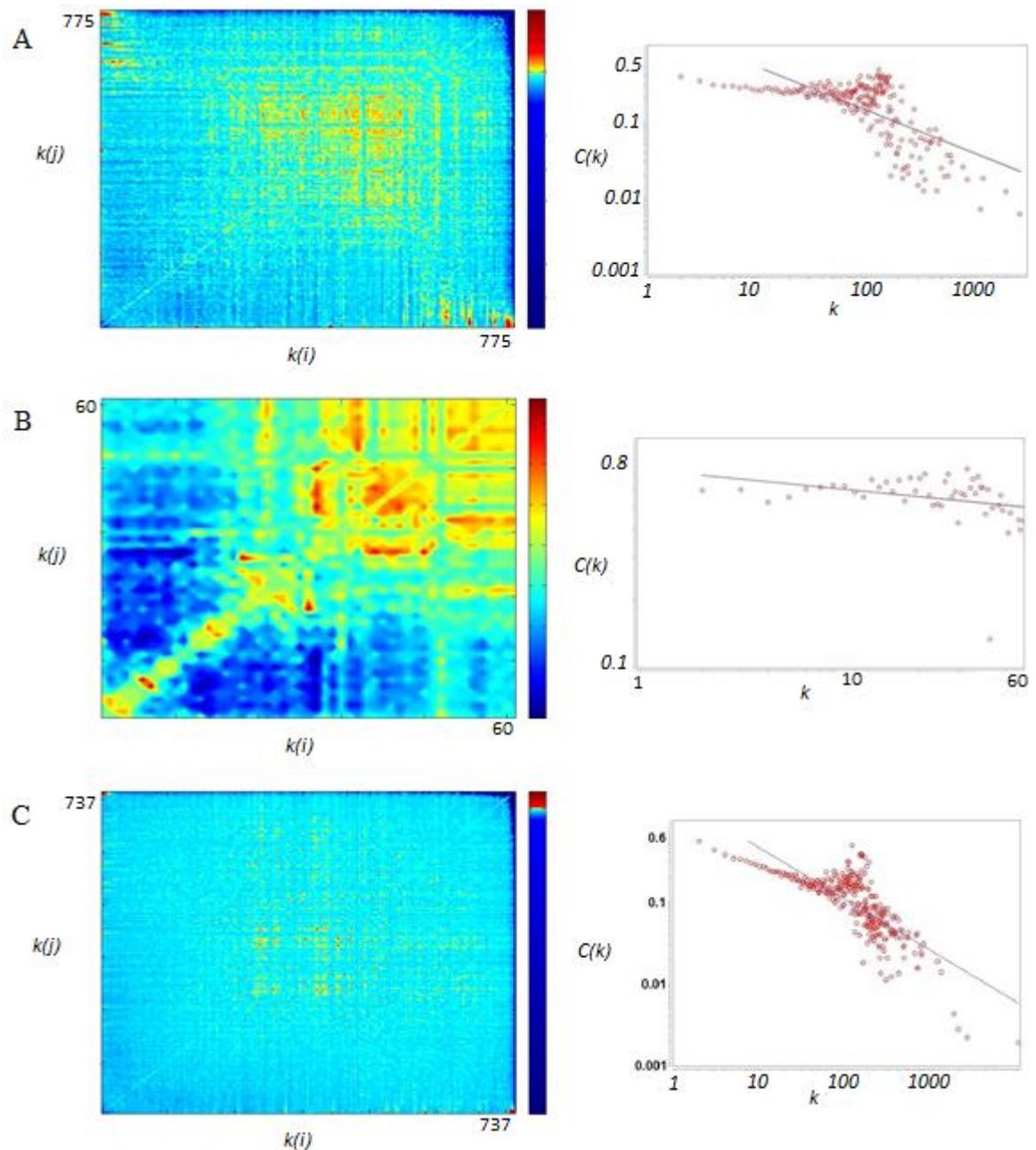


Figure 2.2. Correlation profiles of protein-protein interactions and clustering coefficient variations with degree in (A) Y1, (B) Y2, and (C) H. Corresponding average clustering coefficient variations were fitted to power law line on a logarithmic scale (on the right).

The probability of having a link between two nodes with degrees of  $k(i)$  and  $k(j)$  was compared with the same probability within 100 randomly rewired networks by a z-score. Z-scores that are greater than zero indicate a positive correlation, whereas negative z-scores indicate a negative correlation between connectivities with respect to random networks. Z-scores were used to create heat maps; blue indicates negative and red indicates positive z-scores (Figure 2.2).

Variation of average clustering coefficient with respect to degree and heat-map of z-scores in Y1 network suggests that nodes with a degree of less than 20 are more likely to be linked with highly connected nodes which reflect scale-free nature of the network. There is a strong repulsion between super-hubs whereas moderately linked nodes with degree 84 to 188 have an apparent affinity to each other pointing out the densely connected core of the Y1 network (Figure 2.2a).

Hubs in Y2 network whose degrees are higher than 28 have strong affinity to each other whereas nodes with small degree are less likely to be linked to hubs compared to random networks. Together with apparent correlation in high connectivity it shows that Y2 network has a modular structure with a high inter-modular connectivity (Figure 2.2b).

The assortativity coefficient (Newman, 2002) of the H network was found to be significantly negative ( $r = -0.0717$  with a  $p$ -value  $< 0.01$ ), which implies that the network is disassortative. The variation of the average clustering coefficient with respect to degree and a heat map of z-scores in the H network suggest that nodes with low degree are more likely to be linked to highly connected nodes and nodes with many connections are less likely to be connected to each other, which also indicates the scale-free nature of the network. There is a strong repulsion between super-hubs, whereas moderately linked nodes with a degree of 100 to 200 have an apparent affinity for each other, which points out the densely connected modules (Figure 2.2c). The variation of the clustering coefficient in the H network obeys a power law distribution, which is in accordance with the disassortative nature of the network (Figure 2.2c) and existence of super-hubs.

The correlation between degree and average nearest neighbor degree is another metric that can determine the assortativity of a network (Pastor-Satorras et al., 2001). When average

nearest neighbor degree is plotted as a function of degree the distribution almost perfectly fitted a power law with a negative exponent, which confirms the negative assortativity of the human interactome (Figure 2.3).

Clustering coefficient variation in Y1 and H networks obeys power-law distribution in parallel with the disassortative nature of the networks (Figure 2.2) and existence of super-hubs (Figure 2.1). In Y2 network average clustering coefficient is less responsive to increasing connectivity but it does not imply that the network is not modular because there are no super-hubs (Figure 2.1) and the network is assortative (Figure 2.2).

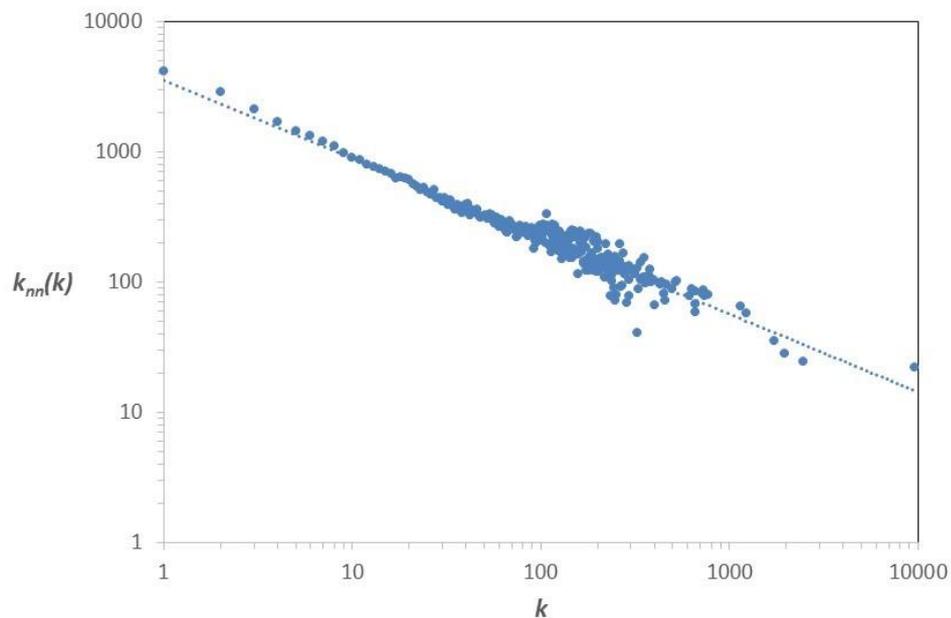


Figure 2.3. Average nearest neighbor degree as a function of degree. Distribution fits to power law with  $R^2=0.934$  (slope of the line is  $-0.597$ ).

### 2.3.3. Network Integrity

Robustness is an important property of protein–protein interaction networks, which is assumed to emerge by natural selection and refers to the ability of networks to maintain their function under perturbation (Waddington, 1942). Efficiency and diameter are the two important topological characteristics used to quantify the robustness of networks (Albert et al., 2000).

All the nodes in Y1, Y2 and H networks were scored using the newly described measure of topological centrality, WSL-EC, and nodes were then removed one by one, either by starting with the highest scoring nodes or randomly. Diameter and efficiency were recalculated after the removal of each node in each of these three biological networks. The results were compared with those when the removal of the nodes was scored by DC, BC, EC and SC (Figure 2.4).

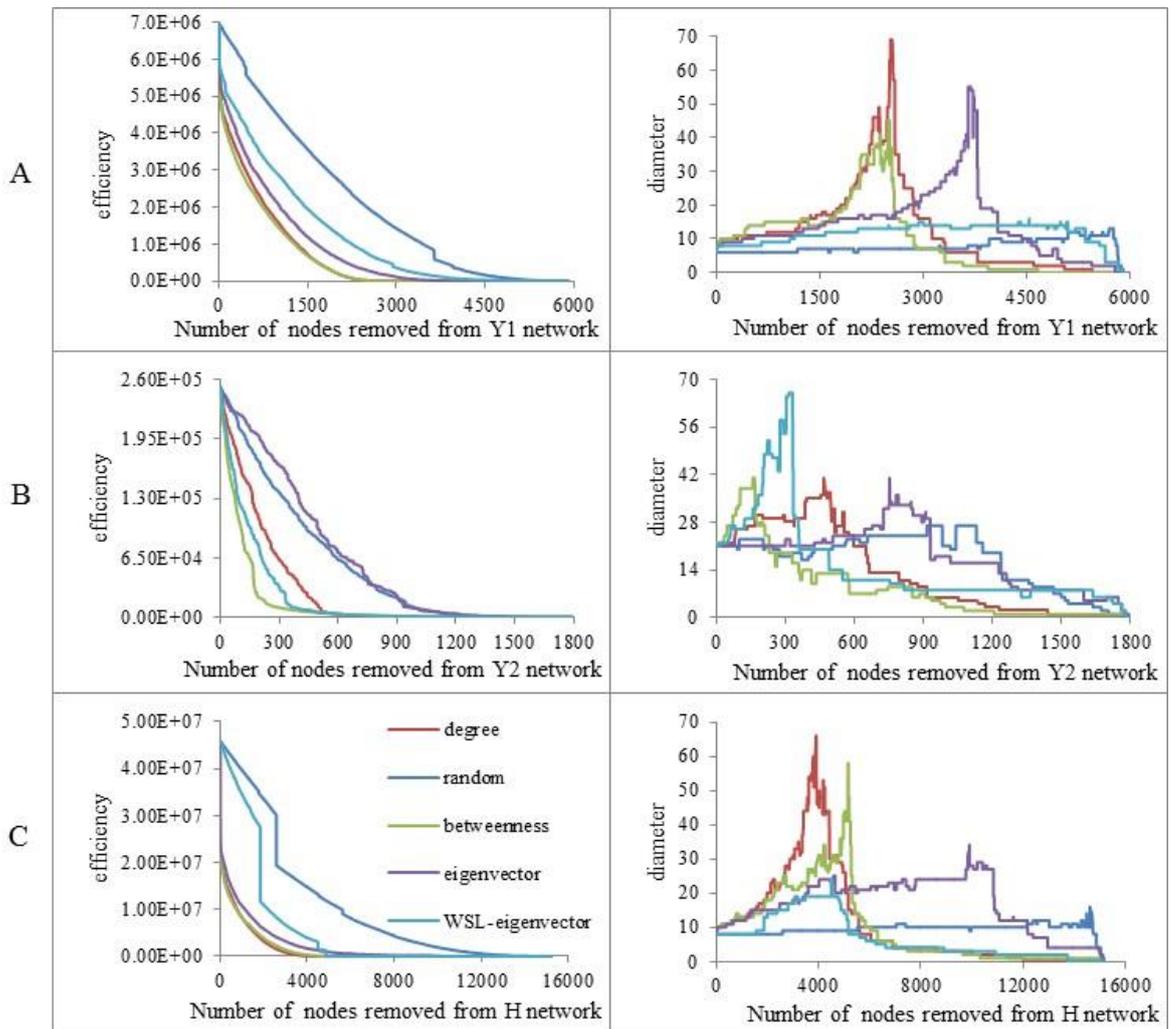


Figure 2.4. Change in efficiency and diameter by random and targeted node removals from (A) Y1, (B) Y2, and (C) H networks.

The drawback of SC revealed itself at the beginning, when the ranking based on SC was found to be identical to the ranking that was found based on EC. The difference between the largest and second largest eigenvalue is more than 53, which means that the weight of the first eigenvector is more than  $e^{53}$  times the weight of the second eigenvector. Hence, the

results for SC will not be presented hereafter as they are identical to those for EC. Removal of the highest-scoring nodes one after the other was found to have more impact in decreasing efficiency when compared with random removal of nodes in all networks. The diameters of the networks were observed to become larger followed by a drastic decrease, which indicated collapse of the network, for targeted removal of the nodes, whereas the diameters of the networks remained unchanged by random removal until these collapsed (Figure 2.4).

The percentage of nodes that has to be removed in a network to decrease the efficiency or diameter to half of its original value was calculated and is presented in Table 2.2. A 50% decrease in efficiency occurred after the removal of 11.7 and 12.2 percent of highest nodes by WSL-EC in Y1 and H networks respectively. This result indicates that the targeted removal of the WSL-EC nodes causes a less severe collapse in this network than the targeted attacks using DC, BC and EC (Table 2.2).

Table 2.2. Percent of nodes removed to attain defined collapse.

Collapse criteria	Network	Random removal	WSL-EC targeted removal	DC targeted removal	BC targeted removal	EC targeted removal
50% drop in Efficiency	Y1	26.2	11.7	5.3	4.0	7.3
	Y2	19.3	5.3	9.2	4.3	22.5
	H	17.1	12.2	0.1	0.1	0.1
50% drop in Diameter	Y1	75.7	42.0	46.9	23.1	70.8
	Y2	96.2	29.6	25.6	33.9	65.2
	H	97.9	43.6	45.4	46.2	83.6

The efficiency drops to its half in Y2 network when 5.3%, 9.2%, 4.3% or 22.5% of nodes were removed by WSL-EC, DC, BC or EC targeted attack, respectively. These results indicate that the removal of the nodes with high WSL-EC score affect network integrity more than the removal of nodes having high DC or EC scores in this network. BC targeted attack was found to cause a fastest disintegration of the network than other targeted attacks in Y2 network (Table 2.2).

Diameter-based criteria, on the other hand, indicate that a targeted attack using WSL-EC causes the fastest disintegration of the H network, which is the opposite of efficiency-based criteria - and the second fastest attack in leading to the same extent of disintegration of Y1 and Y2 networks (Table 2.2).

The drastic difference between these two criteria might possibly be due to the presence of super-hubs in Y1 and H networks and efficiency might be strongly affected by super-hubs. The most dominant super-hub is UBC in the H network, which is connected to more than 63% of all nodes in the network and has by far the highest scores with respect to DC, BC, and EC. However, the WSL-EC ranking of UBC is only 1861, which corresponds to the top 12.2%, which is exactly the percentage of the nodes that need to be removed in order to collapse the H network in a targeted attack using WSL-EC (Table 2.2). Other than UBC, all the top five DC genes and four out of the top five BC and EC nodes are not even in the top 10% of the ranking based on WSL-EC. WSL-EC seems to outperform the other metrics of centrality in identifying nodes that affect network diameter without being affected by the presence of super-hubs in the H network. The presence of super-hubs also explains the dominance of the first principal and consequent convergence of SC to EC.

#### **2.3.4. Hubs**

Four sets that consisted of the top 10% of highest-scoring nodes using DC, BC, EC, and WSL-EC were identified in each network as hub sets. The networks were visualized by Cytoscape in order to investigate and compare the topological distribution of the hubs. The localization of hub sets that were identified using different metrics in the Y1 and H networks were not clearly distinguishable owing to the crowded nature of the networks (Figure 2.5 and Figure 2.6).

The visualization of the four hub sets in the Y2 network indicated that WSL-EC captures the peripheral modules together with the core of the network whereas EC highlights only the core of the network. Hubs that were identified by DC were also found to be localized at densely connected parts of the network. Hubs that were identified by BC and WSL-EC were distributed all over the network (Figure 2.7).

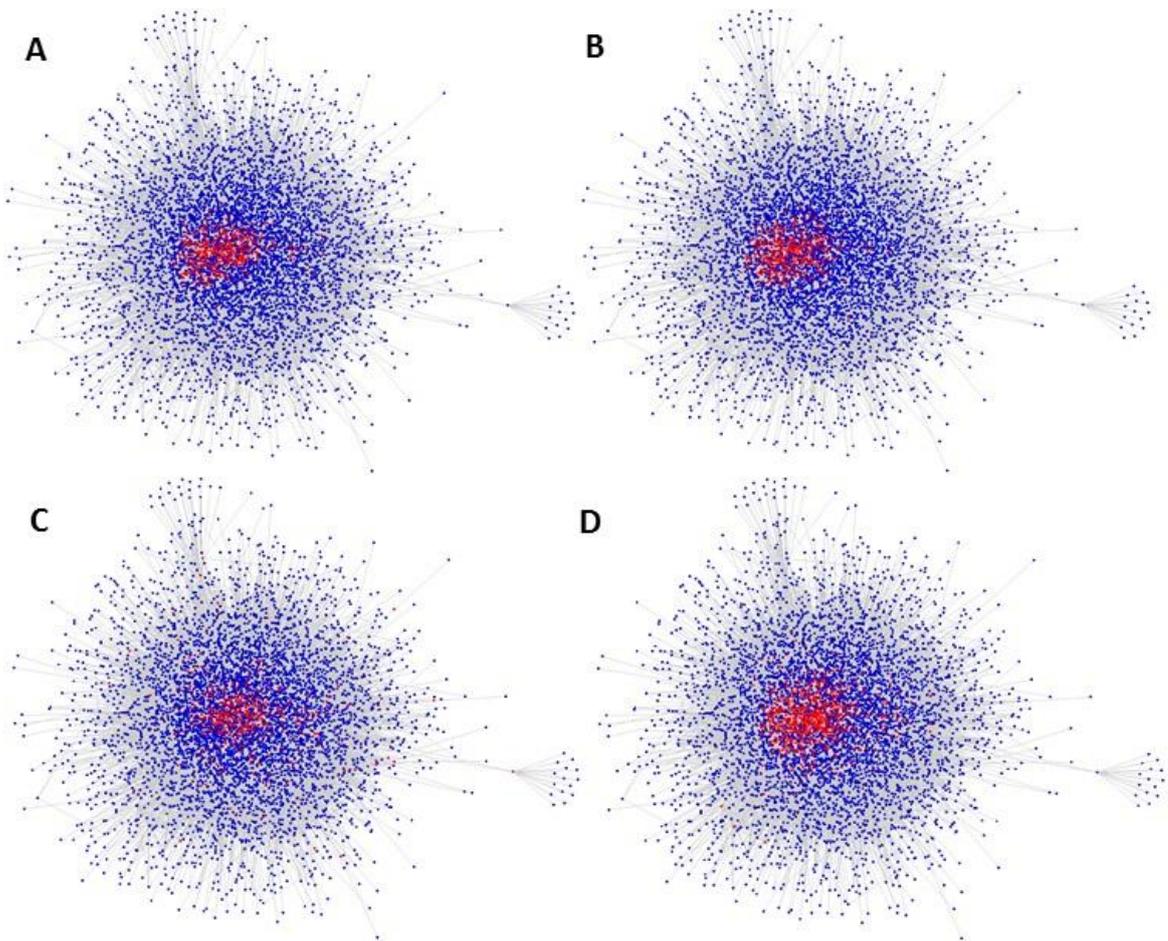


Figure 2.5. Red nodes are hubs defined by (A) DC, (B) EC, (C) BC and (D) WSL-EC in Y1 network.

Because the visualization of large networks may not give satisfactory results, visually detected differences between the dispersions of hubs in the Y2 network were quantified by the topological analysis of hub networks that were constructed using known interactions between the hubs of the Y1 and H networks. Four hub networks, which were constructed using hubs identified by different metrics of centrality, were named as degree central hub network (DC-HN), betweenness central hub network (BC-HN), eigenvector central hub network (EC-HN) and WSL eigenvector central hub network (WSL-EC-HN) for all networks. The network densities of these networks were calculated and compared to quantify the dispersion of the hubs (Table 2.3).

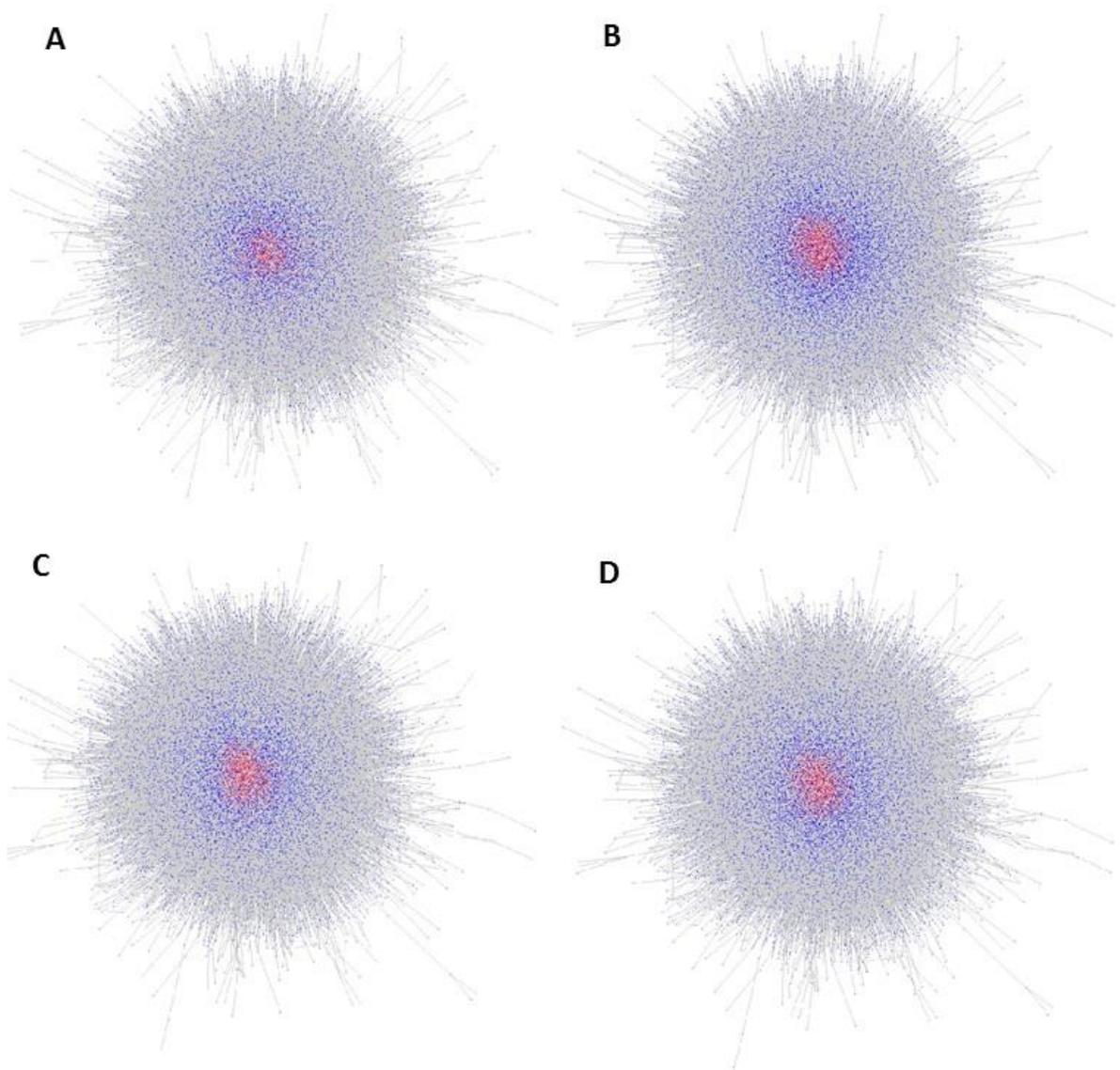


Figure 2.6. Red nodes are hubs defined by (A) BC, (B) EC, (C) DC, (D) WSL-EC in the H network.

The observation that WSL-EC-HN in all three networks has a lower network density than EC-HN and DC-HN indicates that the WSL-EC hubs are more dispersed than the EC and DC hubs, which confirms the visual observation of the hubs in the Y2 network. BC-HN has a lower network density than WSL-EC-HN in Y2 and H networks, which again confirms the visual observation that the BC hubs in Y2 are more dispersed. WSL-EC-HN was found to have the lowest network density in Y1 network indicating almost a similar dispersion of the BC and WSL-EC hubs in all over the Y1 network (Table 2.3).

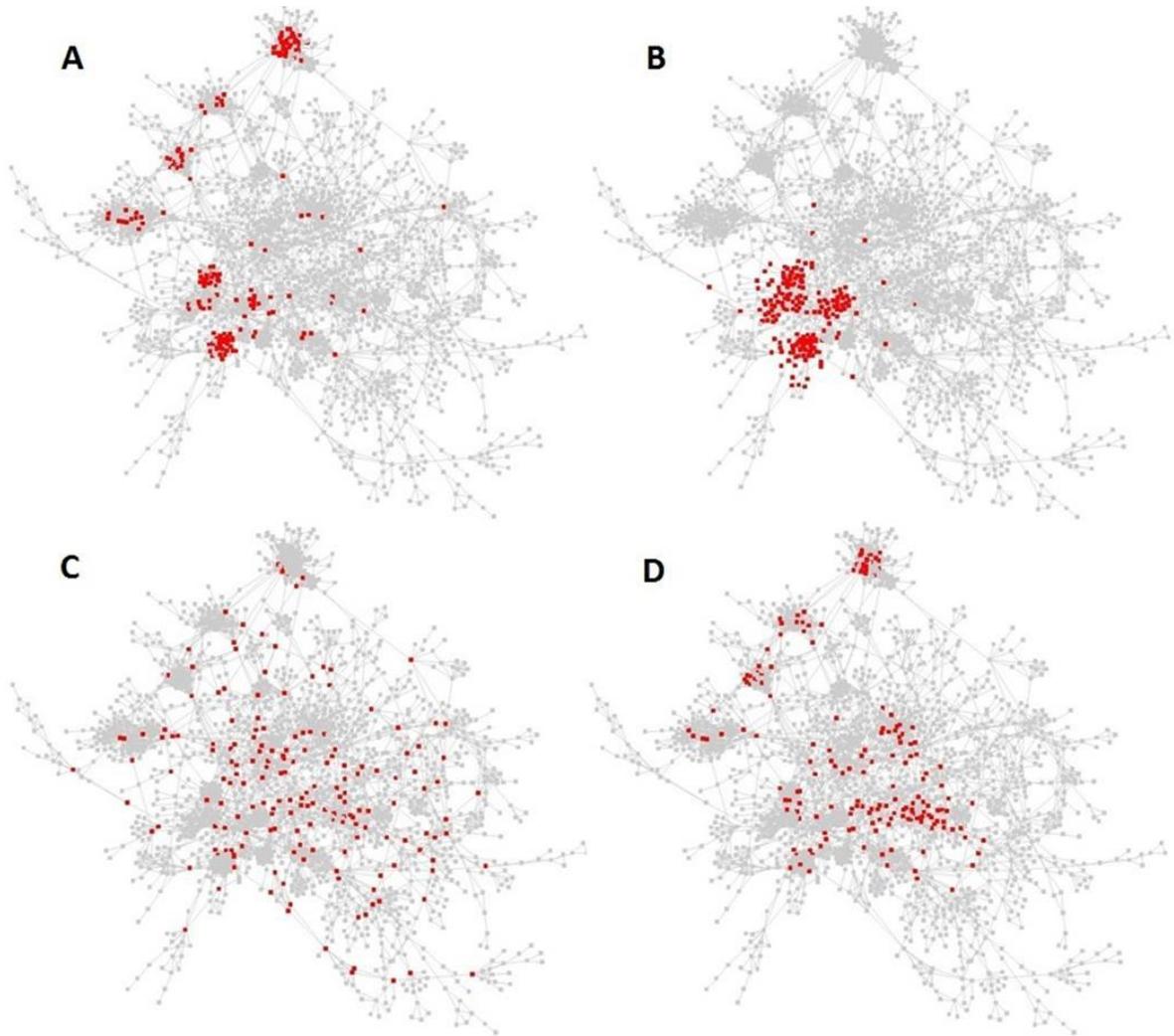


Figure 2.7. Distribution of the hubs defined by (A) DC, (B) EC, (C) BC, (D) WSL-EC in Y2 network.

Table 2.3. Network properties of Hub-Networks.

Network	Hub Network Type	Number of Nodes	Number of Edges	Network Density
Y1	DC-HN	589	16786	0.097
	EC-HN	589	18309	0.104
	BC-HN	589	9512	0.055
	WSL-EC-HN	589	7274	0.042
Y2	DC-HN	179	1657	0.104
	EC-HN	179	1408	0.088
	BC-HN	179	329	0.021
	WSL-EC-HN	179	681	0.043
H	DC-HN	1519	43655	0.038
	EC-HN	1519	43376	0.038
	BC-HN	1519	26106	0.023
	WSL-EC-HN	1519	39625	0.034

### 2.3.5. Biological Investigation of Hubs

The biological role of a node is reported to be strongly related to its topological location within an interaction network, i.e., the functions of interacting neighbors or processes they are involved (Zhu *et al.*, 2007). GO biological process terms that are significantly associated with the four hub sets were also investigated, which consist of the top 10% highest-scoring nodes with the highest centralities with respect to degree centrality (DC), betweenness centrality (BC), eigenvector centrality (EC) and WSL eigenvector centrality (WSL-EC).

A total of 127 hubs were found to be common in all hub sets in Y1 network. These hubs are significantly enriched ( $p\text{-values} \leq 9.76E-4$ ) in GO biological process terms like chromatin organization and its regulation, rRNA processing and regulation of gene expression.

In Y2 network 13 hubs were identified commonly by all centrality measures and were found to be enriched ( $p\text{-values} \leq 1.26E-4$ ) in histone acetylation and proteasome assembly.

In H network almost half of the any hub set is common to all other sets. These 747 human genes have significant ontological associations to a wide variety of GO biological process terms related to ( $p\text{-value} \leq 9.88E-4$ ) communication, response, metabolism, catabolism, localization, development, transcription, cell cycle etc. and regulation of all these processes.

In Y1, 256 hubs were uniquely identified by WSL-EC and found to be significantly enriched ( $p\text{-values} \leq 1.90E-4$ ) with ribosome biogenesis, ribonucleoprotein complex assembly, regulation of translation, nuclear transport, glucose catabolic process, glycolysis, cytoplasmic translation and rRNA processing. 102 hubs specifically described by EC were enriched with biological process terms ( $p\text{-values} \leq 3.71E-4$ ) such as protein-DNA complex subunit organization, chromosome organization, DNA repair, RNA metabolic process and regulation of gene expression. 31 hubs which were selectively identified by DC alone, were found to be significantly enriched ( $p\text{-value} \leq 8.41E-4$ ) with mRNA metabolic process, RNA splicing and RNA polymerase II transcriptional preinitiation complex assembly. When BC

was used as a centrality measure 172 hubs were specifically described and these proteins were found to be significantly associated ( $p\text{-value} \leq 4.00E-4$ ) with cellular response to stimulus, protein phosphorylation, regulation of signal transduction, regulation of DNA metabolic process and regulation of RNA metabolic process (Figure 2.8).

In Y2, 52 hubs were uniquely identified by WSL-EC and found to be significantly enriched ( $p\text{-value} \leq 9.66E-4$ ) with stress granule assembly, membrane budding, transport, regulation of signaling, regulation of cell cycle, regulation of protein localization, growth, signal transduction and cell cycle checkpoint. 77 hubs specifically described by EC were enriched with biological process terms ( $p\text{-value} \leq 6.54E-4$ ) such as protein-DNA complex assembly, regulation of gene expression, RNA biosynthetic process and histone modification. 38 hubs which were selectively identified by DC alone, were found to be significantly enriched ( $p\text{-value} \leq 4.14E-7$ ) with translation, rRNA processing and ribosome biogenesis. When BC was used as a centrality measure 86 hubs were specifically described and these proteins were found to be significantly associated ( $p\text{-value} \leq 2.13E-4$ ) with regulation of biological process, intracellular signal transduction, growth, single-organism process and transport (Figure 2.8).

In the H network, 53 hubs were uniquely identified by WSL-EC and found to be significantly enriched ( $p\text{-value} \leq 8.03E-4$ ) with the establishment of organelle localization and regulation of biosynthetic process. Some 285 hubs were specifically described by EC and were enriched with biological process terms ( $p\text{-value} \leq 1.94E-6$ ) such as gene expression, cellular component organization or biogenesis, organelle organization and mRNA transport. Twenty-six hubs were selectively identified by DC alone and were found to be significantly enriched ( $p\text{-value} \leq 5.49E-5$ ) with regulation of ligase activity, regulation of protein ubiquitination, protein catabolic process and cell cycle. When BC was used as a measure of centrality, 570 hubs were specifically described and these proteins were found to be significantly associated ( $p\text{-value} \leq 6.07E-9$ ) with response to stimulus, biological regulation, multicellular organismal process, signaling, immune system process, developmental process and establishment of localization (Figure 2.8).

The top 10 highest-scoring hub sets that were identified by different metrics of centrality for the H network were compared in order to find individual genes that were

avored by specific metrics of centrality. *RIOK2*, which is a kinase related to the ribosome biogenesis process, is the top central node detected by WSL-EC. The following four nodes *VCAM1*, *ITGA4*, *HSPA8* and *HSPA5* are related to stress response and/or the immune system.

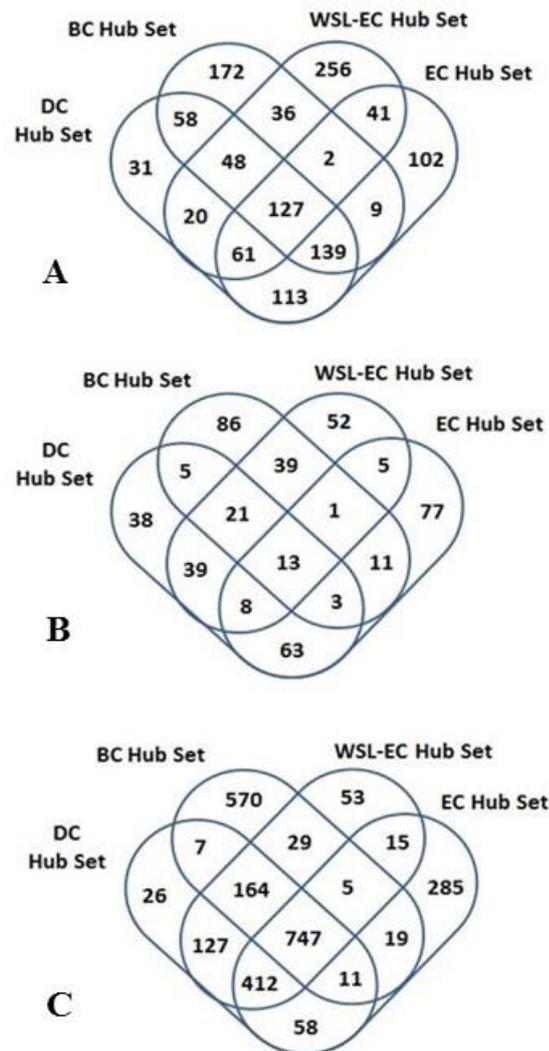


Figure 2.8. Venn Diagram of Degree Central (DC), Eigenvector Central (EC) and WSL Eigenvector Central (WSL-EC) Hub Sets in (A) Y1, (B) Y2, and (C) H Networks.

The top 10 hub sets that were identified by DC, BC and EC were found to be highly overlapping, whereas the top 10 hub set for WSL-EC is completely exclusive. The super-hubs *UBC* and *NRF1* are present in the top two for all DC, BC and EC rankings and *APP*, *ELAVL1*, *SUMO2* and *CUL3* are also present in all three top 10 hub sets. All these nodes are related to ubiquitination processes and/or interactions with RNA.

The top 10 hub sets and pathways that are related to the hubs are tabulated in the Appendix (Tables A2.4-A2.7). The focus of the novel metric seems to be more on immune system process compared with the other metrics of centrality.

### 2.3.6. Centrality and Lethality in Yeast

As soon as genome wide studies emerged the correlation between lethality of a protein and its centrality in the interaction network was detected (Jeong *et al.*, 2001). This correlation has led to the idea that network centrality can be used to estimate essential genes. Therefore in this study, identified top 10% the highest central hub sets in Y1 and Y2 ppi networks were also tested for the enrichment in essential genes (Giaever *et al.*, 2002). Essential gene list in yeast was obtained from Database of Essential Genes (DEG) (Zhang and Lin, 2009). In Y1 network 17.8% of all nodes are essential and ratio of essential nodes in different hub sets varies from 36% (identified by WSL-EC) to 44% (identified by EC) in Y1 (Figure 2.9a).

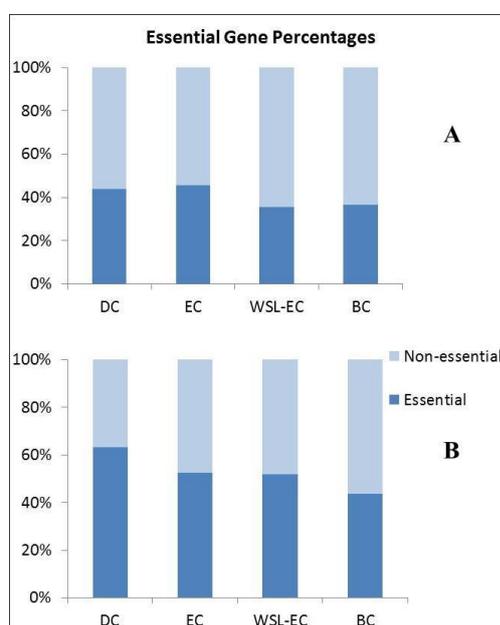


Figure 2.9. Enrichment of essential genes within different hub sets of (A) Y1 and (B) Y2 networks.

A total of 662 genes out of all 1792 Y2 genes (37%) are listed as essential in Database of Essential Genes (DEG) (Zhang and Lin, 2009) and ratio of essential genes determined by different centrality measures were calculated. Hub set identified by DC includes 63% of

essential genes. The hub sets identified by EC, WSL-EC and BC could capture, 53%, 52% and 44% of essential genes present in Y2 respectively (Figure 2.9b).

The percentages of essential proteins within hub sets are subject to alterations with changes in threshold of hub definitions. The correlation between centrality and essentiality was also investigated by a jackknifing method (Holman *et al.*, 2009) based on jackknife resampling technique (Tukey, 1958) without limitations of thresholds. The results imply that essential nodes tend to be more central by means of all four centrality metrics while central nodes detected by DC are more likely to be essential compared to other centrality metrics in Y1 and Y2 networks (Figure 2.10).

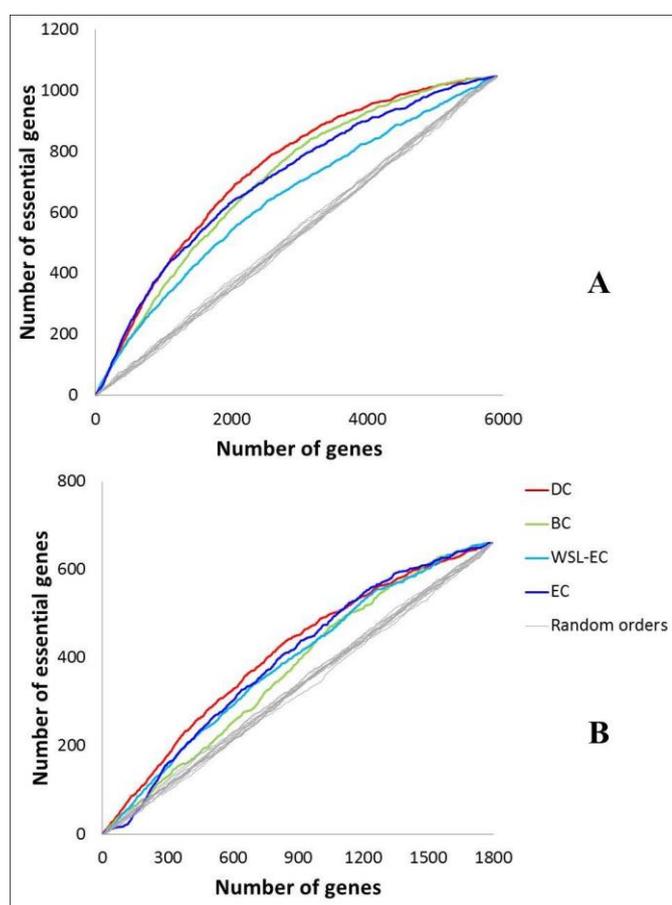


Figure 2.10. Change in number of essential genes detected by centrality ranks and random ranks in (A) Y1 and (B) Y2 networks by jackknifing.

The relative areas under the curve (R-AUC) were calculated for the curves in Figure 2.10. R-AUC was defined as the ratio of the area under the curve (AUC) to the area under

an ideal curve ranking all essential nodes on top (I-AUC) (Table 2.4). As higher R-AUC implies higher correlation with essentiality, DC outperforms the rest in bringing forward the essential nodes in both Y1 and Y2 networks (Table 2.4).

Table 2.4. R-AUC values for DC, BC, WSL-EC, EC and random curves for Y1 and Y2 networks.

	DC	BC	WSL-EC	EC	Random
Y1	0.7035	0.6748	0.6127	0.6736	0.5017 $\pm$ 0.0063
Y2	0.7586	0.6934	0.7207	0.7331	0.6156 $\pm$ 0.0078

All centrality based ranks correlates well with the ideal rank compared to random ranks implying that essential nodes have higher topological centralities than non-essential nodes. Correlation between WSL-EC based centrality and lethality is higher in Y2 network compared to Y1 network possibly due to the differences in the network architectures.

The low correlation between centrality and essentiality by WSL-EC may be explained by the locations of high scoring nodes which are dispersed in all over the network. DC or EC scored nodes were visualized at the core where possibly subunits of the protein complexes were closely localized (Figure 2.7).

Recent studies report that not the centrality but the neighborhood of a node which is densely connected, biologically interrelated and enriched in essential proteins is the reason for essentiality of a node (Zotenko *et al.*, 2008). It was also reported that larger protein complexes are more likely to be essential (Wang *et al.*, 2009; Song and Singh, 2013).

We also investigated the, enrichments of protein complexes in essential genes. Previously reported 430 protein complexes in yeast (Baryshnikova *et al.*, 2010) were used as a template. In Y1 network 1845 proteins were found to be subunits of 429 protein complexes, out of which 629 proteins were essential. When the top high scoring 1845 proteins were investigated, 638, 505, 605 and 576 essential proteins were identified out of 1845 highest scoring nodes with DC, WSL-EC, EC and BC, respectively.

In Y2 network 916 proteins were detected as subunits of 244 protein complexes. 411 protein complex subunits were detected as essential. When the top high scoring 916 proteins were investigated with respect to DC, WSL-EC, EC and BC the number of essential nodes are 457, 416, 441 and 401, respectively. These findings provided further evidence that the essentiality of a gene is not due to its centrality but due to the biological role of the densely connected neighborhood of the gene

### **2.3.7. Biological and Topological Centrality in the Human Interactome**

It has been reported that the genes involved in cancer, aging and infectious disorders are also topologically central (Milenković *et al.*, 2011). Disease-related genes, genes involved in immune system process and genes related to autoimmune diseases also have biological importance. The topologically central hub sets that were identified by the four measures of centrality in the H network were investigated in terms of the enrichments in HIV-1 interacting proteins, ageing-related genes, cancer-related genes, pathogen-interacting (PI) proteins, disease-related genes, genes involved in immune system process and genes related to autoimmune diseases in order to shed light on biological differences in the metrics of centrality.

The correlation between biological and topological centralities was investigated by a jackknife method (Holman *et al.*, 2009) based on a jackknife resampling technique (Tukey, 1958) without limitations of thresholds (Figure 2.11). The relative areas under the curve (R-AUC) were calculated for the curves in Figure 2.11. R-AUC was defined as the ratio of the area under the curve (AUC) to the area under an ideal curve that ranks all biologically significant nodes on top (I-AUC) (Table 2.5). As a higher R-AUC implies a higher correlation with biological significance, analyses of R-AUC indicated that WSL-EC-based rankings outperform DC, BC, EC or SC-based rankings in identifying all biologically central node sets. The statistical significance of the differences was assessed by permutation tests ( $p$ -value  $< 0.001$  for all curves).

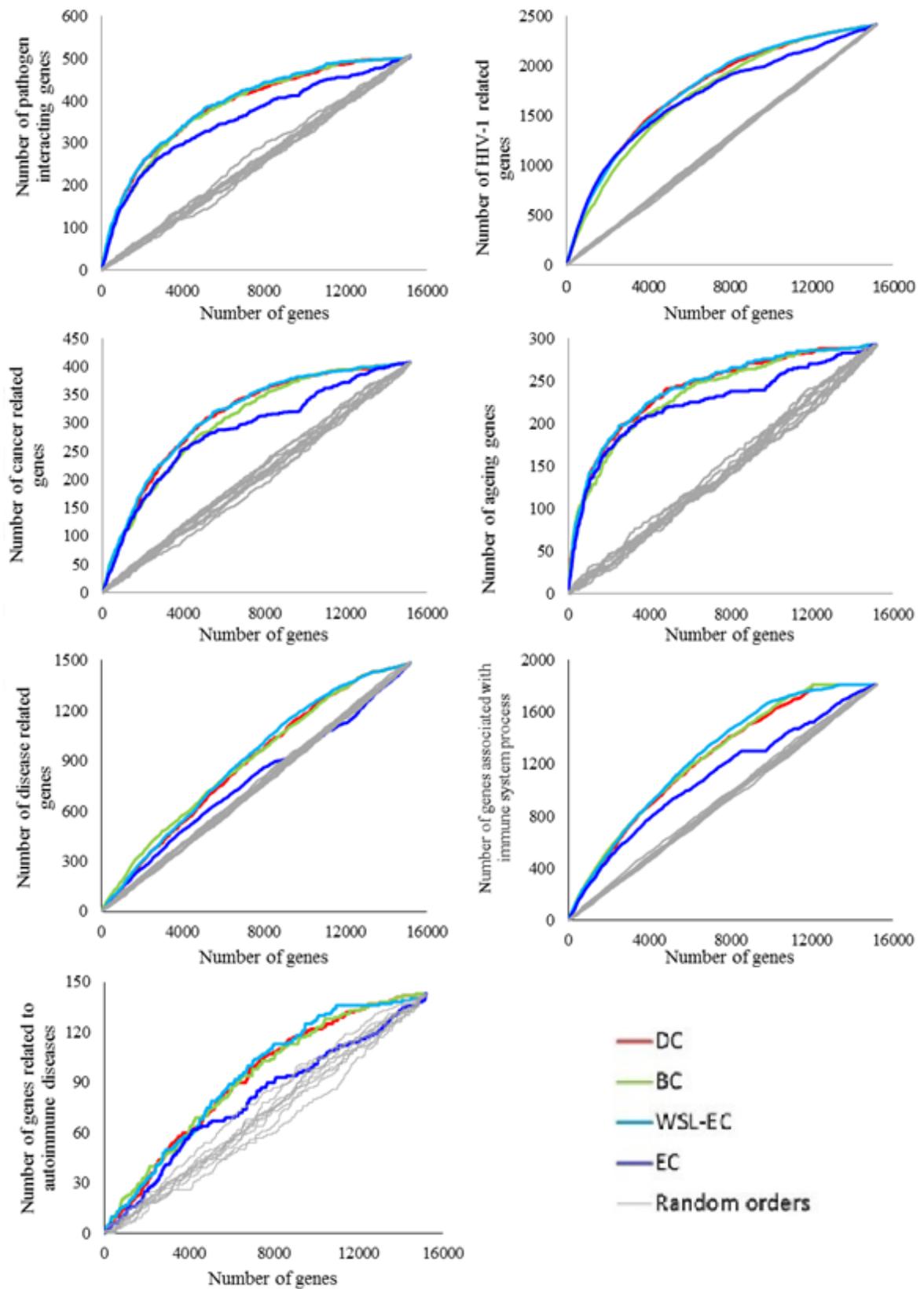


Figure 2.11. Change in number of biologically significant genes detected by centrality ranks and random ranks in the H network.

GO biological process term enrichments were used to distinguish subsets of the biologically central nodes with higher rankings with respect to each metric of centrality. All subsets of biologically central node sets that were defined by seven criteria (pathogen-interacting, HIV-1-interacting, cancer, ageing, and disease-related nodes and genes involved in immune system process or related to autoimmune diseases) and had the highest rankings in the WSL-EC order were found to be enriched in stress response-related terms, immune system-related terms, transcriptional terms and/or kinase activity-related terms.

Table 2.5. R-AUC values for DC, BC, WSL-EC, EC and random curves for the H network.

	DC	BC	WSL-EC	EC	Random
Pathogen Interacting Genes	77.91	77.54	<b>78.73</b>	71.57	50.63 ± 1.22
HIV-1 Related Genes	79.30	76.92	<b>79.45</b>	76.20	54.26 ± 0.51
Cancer Related Genes	77.54	75.42	<b>77.96</b>	71.29	50.92 ± 1.41
Ageing Related Genes	83.05	81.18	<b>83.42</b>	77.37	49.43 ± 1.18
Disease Related Genes	62.49	63.49	<b>63.58</b>	54.79	52.32 ± 0.56
Genes Associated with Immune System Process	72.67	73.33	<b>74.16</b>	64.65	53.53 ± 0.71
Genes Related to Autoimmune Diseases	64.69	64.84	<b>65.75</b>	56.11	49.96 ± 2.84

The smallest subsets for all criteria were found to consist of nodes that were favored by DC. The subsets are either not significantly associated with any GO term or enriched in transcriptional, apoptotic and cell cycle processes.

The widest range of GO terms was determined for the subsets of nodes with high betweenness for all four criteria of biological centrality. External processes, such as exocytosis, endocytosis, cell motility, cell adhesion, cell migration, and cell–cell communication, or processes related to differentiation, such as regulation of neurogenesis and embryonic morphogenesis or immune system-related GO terms, are some of these. The nodes with the highest order in the EC ranking have the narrowest range of GO enrichments as they are found to be associated with only translational and carbon central metabolism-related terms.

## 2.4. Discussion

In the present study, a novel global metric of centrality, weighted sum of loads eigenvector centrality (WSL-EC), counting all eigenvectors was proposed. The performance of WSL-EC in the identification of topologically more important nodes that contribute to the integrity of a network and in capturing essential or biologically central nodes was tested in three biological networks and compared with the performances of four other commonly used metrics of centrality, DC, BC, EC and SC.

Topological analysis of these networks indicated a similar network architecture for yeast global protein-protein interaction network (Y1) and human global protein-protein interaction network (H) which is slightly different than that of the functional subnetwork of yeast protein-protein network (Y2).

Topological analysis of the networks also indicated that the global human protein-protein interaction network (H) has a dissortative and modular architecture. WSL-EC outperformed DC, BC, EC and SC in identifying nodes that affect network robustness in the human interactome. The topological distributions of hubs in the networks were found to be different for hub sets that were identified by different metrics of centrality. Hubs that were identified by BC and WSL-EC were distributed all over the network, whereas hubs that were identified by EC and DC were localized at densely connected parts of the networks.

It was noted that different measures of centrality could specifically capture sets of hubs involved in different biological processes.

WSL-EC and other centrality metrics outperformed in capturing essential genes in Y2 network compared to Y1 network possibly due to the differences in the network architectures and the reliability of the data used in the construction of Y2.

WSL-EC was found to outperform in capturing biologically central nodes, such as pathogen-interacting, HIV-1, cancer, ageing, and disease-related genes and genes, involved in immune system process and related to autoimmune diseases in the human interactome compared with DC, BC, EC or SC. The choice of metric of centrality is crucial, as different

metrics focus on different topologies and these topological differences correspond to different biological roles.

Hubs with off-the-scale connectivity (super-hubs) create a strong bias in topological centrality for DC, BC, EC and SC, whereas WSL-EC does not seem to be affected by the presence of super-hubs.

WSL-EC is an easy-to-implement metric, which does not require a special code or complicated computations. It is promising in the respect that it can be utilized by diverse researchers.

## **2.5. Future Prospects**

The novel metric of centrality, WSL-EC, displays substantial biological relevance and further studies will be required to test the performance of this novel metric of centrality in complex biological networks to reveal the correlation between topology and biological importance.

The performance of the metric in directed or weighted biological networks can also be studied. Furthermore, integration of other data sets with protein–protein interaction networks should be investigated to improve its performance across different network architectures.

### **3. INVESTIGATION OF DYNAMIC RESPONSE OF YEAST CELLS TO A CHEMOTHERAPEUTIC AGENT: DOXORUBICIN**

Dynamic changes in yeast cells were investigated as a response to an impulse of widely used chemotherapeutic agent doxorubicin. In this chapter transcriptional re-arrangements that take place after introduction of chemical stress were analyzed by well-established methods and some modified and developed versions of these methods. Temporal organization and prominent patterns of the global transcriptional response were identified by decomposition and clustering techniques. Integrative approaches were also utilized in order to assess depth in the response regarding physical interactions and regulatory associations. Differentially expressed gene sets were investigated and novel extensions were provided to the methods which extracts differentially expressed genes. Resulting subsets were reanalyzed comparatively.

#### **3.1. Background**

Chemotherapeutic agents in cancer treatment lead to severe side effects and chemotoxicity in long term and cancer cells may develop resistance to these drugs in time (Cheung-Ong *et al.*, 2013). It is crucial to fully understand cellular mechanisms affected by these agents. Doxorubicin is a powerful chemotherapeutic drug that can be used wide range of cancer types (Di Marco *et al.*, 1969). It is proposed that doxorubicin interacts with DNA through intercalation like other anthracycline antibiotics leading to DNA damage (Fornari *et al.*, 1994; Anders *et al.*, 2013). It is believed that during the transcription doxorubicin inhibits topoisomerase II and prevents formation of double helix (Pommier and Marchand 2012). Despite the considerable efforts, effect mechanisms of doxorubicin have not been completely enlightened.

*Saccharomyces cerevisiae* is well defined and easy to manipulate model organism in order to investigate cellular response to doxorubicin in a dynamic context. The effect of doxorubicin dosage on vitality of *S.cerevisiae* and sensitivity of haploid deletion mutants of

*S.cerevisiae* to the agent have already been investigated (Xia *et al.*, 2007) but there is no work in the literature investigating dynamics of doxorubicin response of *S.cerevisiae*.

Within the era of rapidly developing experimental technologies now it is possible to quantify cellular responses to any stress source in time and in omics level and there is great efforts to analyze dynamic data efficiently but this challenge is still largely unfulfilled.

The most common approaches are hierarchical clustering and principal component analysis not only for static transcriptomic datasets but also for dynamic datasets. Hierarchical clustering is in the form of linkage tree based on pairwise similarities and has been commonly used for microarray data analysis all along (Eisen *et al.*, 1998). PCA might be the most common methodology elucidating dominating expression trends; it can summarize microarray data by dimension reduction (Raychaudhuri *et al.*, 2000).

Self-organizing maps (SOMs) are ideal for exploratory data analysis; they can extract predefined number of clusters which reflect prominent expression profiles (Tamayo *et al.*, 1999).

Weighted gene co-expression network analysis (WGCNA) is a popular unsupervised approach in life sciences community (Zhang and Horvath 2005). WGCNA proposes a framework for soft thresholding in order to weight pairwise expression correlations to create a weighted co-expression network then defines node dissimilarities based on co-expression weights and modulates transcriptome based on the topological overlap of the dissimilarities. Meta-analysis of the resulting modules by eigengene networks is also possible (Langfelder and Horvath 2007).

It has been shown that integrating multi-omics data can unveil additional information about biological systems (de Keersmaecker *et al.*, 2006). Interactome or regulome can be integrated with transcriptome to elucidate the results. NP analysis assumes existence of global subnetworks specific to perturbations and uses positive and negative correlations of dynamic transcription profiles in order to reduce protein-protein interaction network into an active subnetwork (Xia *et al.*, 2006). Dynamic regulation of cells in response to any perturbation can be investigated by integrating resulting transcriptome with transcriptional

factor (TF) -DNA interactions by using dynamic regulatory event miner (DREM) which is tailored for time series expression experiments (Schulz *et al.*, 2012).

Another common approach is to reduce datasets to significant subsets by identifying differentially expressed genes. Typically fold change can be used for this purpose for static datasets whereas identification of the differentially expressed genes out of time series transcriptome data has been problematic. Significance analysis can be carried out by a methodology specifically designed for dynamic datasets which is called as EDGE (Storey *et al.*, 2005). The idea behind the method is to identify genes whose expression profiles significantly fluctuate which ensure eliminating low abundant transcripts.

In this study chemostat experiments were implemented in a fermenter and transcriptome data were collected after a doxorubicin pulse. The dataset were analyzed by well-established clustering methods and integrative approaches. Differentially expressed subset was also identified. Existing methodologies were extended or modified wherever it is needed.

## 3.2. Materials and Methods

### 3.2.1. Strain, Growth Conditions, Pulse Injection and Sampling

Homozygous *ho* $\Delta$ /*ho* $\Delta$  strain of *Saccharomyces cerevisiae* diploid BY4743 (*MATa/MAT $\Delta$  his3 $\Delta$ 1/his3 $\Delta$ 1 leu2 $\Delta$ 0/leu2 $\Delta$ 0 lys2 $\Delta$ 0/+ met15 $\Delta$ 0/+ ura3 $\Delta$ 0/ura3 $\Delta$ 0) was used in this study. The precultures were incubated overnight in F1 media (Baganz *et al.*, 1997) at 30 °C and 180 rpm in an orbital shaker.*

Chemostat experiments were carried out in 2 L B-Braun Biostat B fermenters with 1.5 L working volume under aerobic conditions in F1 media at a dilution rate of 0.1h<sup>-1</sup>. The fermentation temperature and pH were controlled at 30 °C and pH 5.5, respectively. Fermenters were stirred at 800 rpm which, together with constant air flow at a rate of 0.1 vvm, provided dissolved oxygen at  $\geq 80\%$  dO<sub>2</sub> saturation at all times during cultivation.

During the chemostat experiment, the oxygen and carbon dioxide concentrations in the off-gas,  $dO_2$ , pH, temperature, and the added amounts of base + antifoam were monitored online. The carbon dioxide and oxygen volume fractions in dried off-gas were measured online with a combined carbon dioxide and oxygen gas sensor (BlueSens, Herten Germany).

Samples for transcriptome were taken after spending 5 residence times at steady state. Samples were collected at 0 min (before the doxorubicin pulse) and then at 1, 5, 10, 15, 20, 25, 30, 60, 90, 120 and 180 minutes after the pulse. Doxorubicin pulse was prepared to attain a final concentration of  $20\mu\text{M}$  by solving 17.4 mg doxorubicin hydrochloride (Sigma, Cat. No: D1515) in water.

### **3.2.2. Sampling, RNA Isolation and Preprocessing the Data for Transcriptome Analysis**

For biomass determination, samples at steady-state were collected and washed twice with distilled water followed by centrifugation (8000 rpm, 6 min). The cell dry weight was obtained gravimetrically.

For transcriptome analysis, a culture sample of 4 ml at each time point was withdrawn from the bioreactor and immediately frozen in liquid nitrogen and stored at  $80^\circ\text{C}$  until RNA isolation. Total RNA was isolated in a robotic workstation, QIAcube (Qiagen, USA) in accordance with enzymatic lysis protocol which is described by Qiagen RNeasy mini kit (Cat no: 74106). The quantity and quality ( $A_{260}/A_{280}$ ) of the RNA were assessed by using a micro-volume UV-vis spectrophotometer (NanoDrop ND-1000, Thermo Fisher Scientific Inc., USA) and a microfluidics-based platform Bioanalyser 2100 (Agilent Technologies). cDNA was synthesized, and double-stranded cDNA was synthesized from ca. 100ng of total RNA as described in the Affymetrix GeneChip<sup>®</sup> Expression Analysis Technical Manual, using appropriate kits. cDNA was quantified using Nanodrop spectrophotometer.  $5\mu\text{g}$  of aRNA was loaded onto Affymetrix Yeast2 arrays. The chips were then loaded into a fluidics station for washing and staining using Affymetrix Comman Console<sup>®</sup> Software (AGCC) 3.0.1 Fluidics Control Module with Mini\_euk2v3. Lastly, the chips were loaded onto the Agilent GeneArray scanner 3000.

### 3.2.3. Transcriptome Data Analysis

The raw data were processed with dChip software for outliers at the array level (Li and Wong, 2001). RMA Express software was used to normalize and log<sub>2</sub> transform the data (Bolstad *et al.*, 2003).

3.2.3.1. Identification of differentially expressed genes. Significance analysis has been carried out by EDGE which is specifically designed for dynamic datasets. An R software package with the same name “edge” (Storey *et al.*, 2015) was used for the application of the methodology. Optimal discovery procedure (odp) option for statistics was used (Storey *et al.*, 2007).

3.2.3.2. Principal Component Analysis. PCA has been carried out by GeneCluster 3.0 software package (de Hoon *et al.*, 2004).

3.2.3.3. Self-Organizing Maps. GeneCluster 2.1.7, a software package (Reich *et al.*, 2004), was used to identify SOMs of the time series response of yeast cells to a doxorubicin pulse.

There are algorithms to estimate optimum number of clusters to be set for SOMs. Here in this study in order to manage clusters manually the highest plausible number was determined. Typically the number of clusters can be set from 2 to  $\sqrt{N}$  (where  $N$  is the number of nodes) (Vesanto and Alhoniemi 2000). Thus, SOMs of 6 rows and 6 columns (36 clusters  $> \sqrt{1150}$ ) were set.

3.2.3.4. Hierarchical Clustering. A software package, Hierarchical clustering explorer (HCE 3.0), was used for the analysis (Seo and Shneiderman 2004).

3.2.3.5. Weighted Gene Co-expression Network Analysis (WGCNA). The effect of doxorubicin pulse on yeast chemostat culture was analyzed with WGCNA in R environment (Langfelder and Horvath 2008).

3.2.3.6. NP Analysis. The initial global interaction network to a global active subnetwork was reduced by NP analysis as described by Xia et al, in 2006 (Xia et al, 2006) where NP stands for negative and positive correlations.

NP analysis relies on expression correlations and anticorrelations between pairs of genes which are measured by Pearson correlation coefficient (PCC). PCC thresholds for correlation and anti-correlation were decided in each application separately.

The interactions between nodes that are neither correlated nor not anti-correlated as well as the nodes with no interaction were eliminated from the PPI network. The remaining network is called as “NP network” and assumed to be a global active subnetwork of the reference global PPI network.

The NP network was divided into two largest possible modules displaying opposite expression profiles were extracted using Hierarchical Clustering Explorer (HCE) software. The largest anti-correlated clusters, so that the clusters that have less than 1% intra-cluster anti-correlated interactions, were manually dissected.

3.2.3.7. Dynamic Regulatory Event Miner (DREM). The model DREM was used to integrate dynamic transcriptome data with static TF-DNA interaction data. DREM is built on Input-Output Hidden Markov Model (IOHMM) (Bengio and Frasconi, 1995). The details of the likelihood function and model learning were explained by Ernst and colleagues (Ernst *et al.*, 2007).

Once the probabilistic tree representing the map of regulation is constructed, DREM identifies master regulators by scoring transcriptional factors for each branch by using hypergeometric distribution. The score of  $TF_x$  for the path  $A$  is calculated as described in Equation 4.1.

$$score(TF_x, A) = \sum_{i=C_A}^{\min(C_S, n_A)} \frac{\binom{C_S}{i} \binom{n_S - C_S}{n_A - i}}{\binom{n_S}{n_A}} \quad (4.1)$$

where,  $n_S$  is total number of genes before the split,

$n_A$  is number of genes in the branch  $A$ ,

$C_S$  is total number of genes regulated by  $TF_x$  before the split,

$C_A$  is number of genes in the branch  $A$  which are regulated by  $TF_x$

The regulatory map is constructed by using TF-DNA associations that is why this score is not an actual *p-value* (Ernst *et al.*, 2007).

TF cut-off score of *0.001* was used to identify master regulators.

3.2.3.8. Functional Annotations. Gene ontology biological process term and pathway enrichment analysis were carried out by The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (Huang *et al.*, 2009). Enrichments whose *p-values* were less than 0.01 were considered as significant.

Selected GO terms lists were prepared in order to cover all identified terms based on a term-tree generated by using QuickGO browser (Binns *et al.*, 2009).

### 3.3. Results

Doxorubicin was injected to a yeast chemostat culture and both short-term and long-term response were monitored by sampling at non-uniform time intervals, ranging from minutes to hours in transcriptomic landscape. Temporal organization of the global transcriptional response of yeast cells to a doxorubicin pulse was analyzed by using different approaches. Time points were clustered and then global transcriptome was analyzed by hierarchical clustering and principal component analysis as a first step. Self-organizing maps (SOMs) and weighted gene co-expression network analysis (WGCNA) were used after filtering whole transcriptome into significant sub-datasets. Integrative approaches were also used to analyze the transcriptome data. Transcriptome data was integrated with physical interaction network by NP analysis approach and with protein-DNA interaction network through dynamic regulatory event miner (DREM) approach to identify key transcriptional factors which regulate the dynamic response.

Secondly differentially expressed genes were identified by using EDGE method which is specifically designed for time series experiments. In order to ensure high number of true positives a pipeline was proposed to extend differentially expressed genes.

Finally NP analysis, DREM and WGCNA methods were applied to the subset of differentially expressed genes identified by the extended EDGE and compared the results with the previous ones. Some modifications on the methods were also proposed and their results were discussed.

### 3.3.1. Temporal organization of Global Transcriptional Response

Doxorubicin was injected to a yeast chemostat culture and both short-term and long-term responses were monitored by sampling at non-uniform time intervals, ranging from minutes to hours in transcriptomic landscape. Hierarchical clustering of time points of the transcriptome in response to a doxorubicin pulse revealed that the response was clustered into 3 groups (Figure 3.1). The global response of yeast cells to the impulse like addition of the chemical indicated that the steady state data and the transcriptomic response within first 1 minute were clustered together and the data taken at 5<sup>th</sup> minute was connected to this first cluster.



Figure 3.1. Hierarchical clustering of time points

This observation indicated that the cells have started to reorganize their response one minute after the injection of the pulse. The data taken at 10<sup>th</sup> up to 30<sup>th</sup> minutes form another cluster and finally last 4 time points were grouped in a third cluster. These clusterings indicated that the response of yeast cells to doxorubicin may be classified into short-term, mid-term and long-term responses.

### 3.3.2. Analysis of Global Transcriptional Response

Global transcriptional response of the yeast cells was analyzed using hierarchical clustering and dynamic expression profiles were clustered using PCA, SOMs and WGCNA. The significantly associated biological process terms with each cluster were identified and the affected pathways were determined.

3.3.2.1. Hierarchical Clustering of Global Response. Hierarchical clustering is one of the most common method to analyze microarray data. Hierarchical clustering approach builds a tree of genes starting from the pair with the closest correlation. The global dynamic transcriptional response of *Saccharomyces cerevisiae* cells to a doxorubicin pulse consisting of the expression levels 5655 genes at each sampling point was hierarchically clustered using HCE 3.0. Minimum similarity was set to 0.8 as cutoff value by analyzing the variation of minimum similarity with resulting number of clusters and with the size of the largest cluster (Figure 3.2).

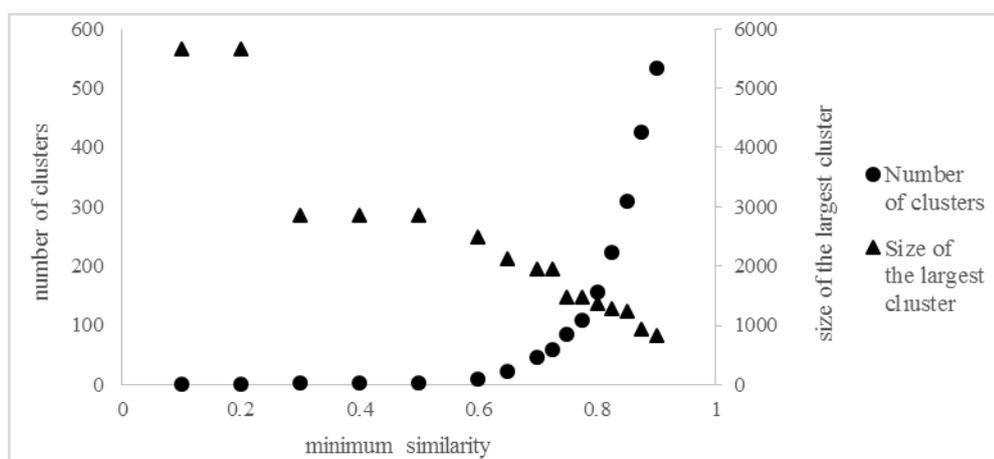


Figure 3.2. Variation of number of clusters and size of the largest cluster with minimum similarity.

Hierarchical clustering of the expression levels of the genes was presented in Figure 3.3. The transcriptomic data was clustered to 156 clusters. The largest cluster consists of 1361 genes (A) and the second largest one consists of 1204 genes (B) (Figure 3.3).

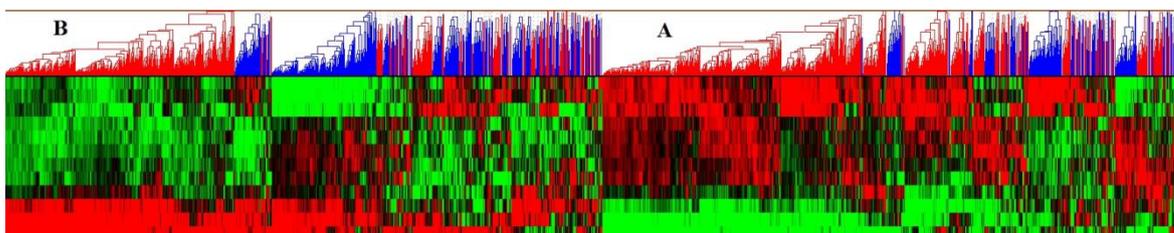


Figure 3.3. Hierarchical clustering dendrogram of dynamic response to doxorubicin in yeast cells. (A) is the largest cluster and (B) is the second largest cluster.

The variation of the average expression profiles of the genes in clusters A and B with time showed that the genes in cluster A and B have an opposite tendency in the direction of change with time. The genes in cluster A have a down-regulating tendency whereas the genes in cluster B have a tendency to be up-regulated with time (Figure 3.4).

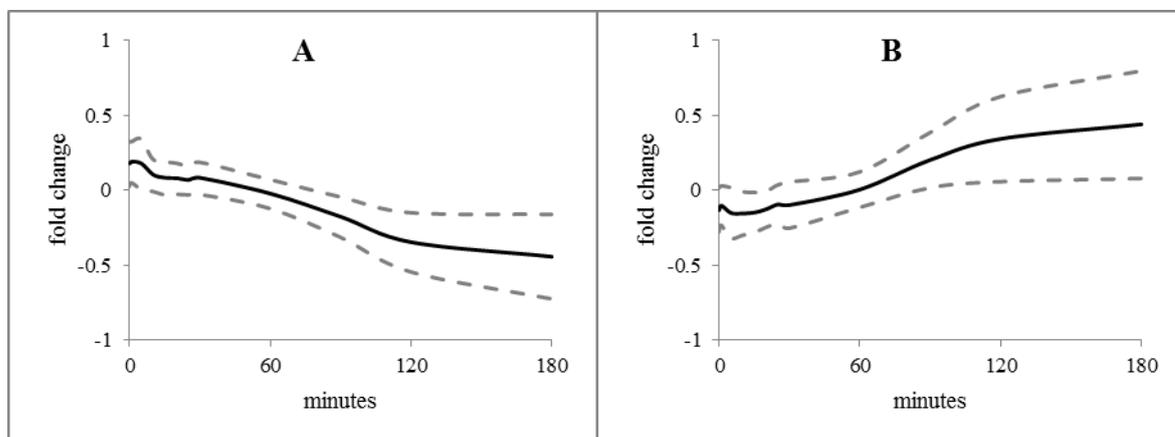


Figure 3.4. The variation of the average normalized expression profiles of the genes in A and B clusters (A and B respectively). Dashed lines represents standard deviation.

**3.3.2.2. Principal component analysis (PCA).** Principal component analysis is a powerful tool in order to obtain the most prominent trend throughout the dataset. Here PCA was applied to the transcriptome data to get the most common expression profiles by eigenvector decomposition.

Comparative analysis of the whole dataset and of a subset of genes (with high standard deviation  $> 0.25$ ) by PCA indicated that the selection of a sub-set of transcription profiles did not improve the outcomes of the analysis. Therefore, in order to avoid losing the data, analysis were carried on with the whole dataset.

The top 4 principal components resulting from PCA represents more than 80% of the total variation in the dataset (Figure 3.5).

The first principal (PC1) of the transcriptome data (PC1) covers 61% of the total variance. 1466 genes, which were positively correlated with PC1 ( $PCC > 0.7$ ), display an up-regulation after 30 minutes. 1455 genes negatively correlated with the first principal ( $PCC < 0.7$ ) which implies a down-regulation after 30 minutes.

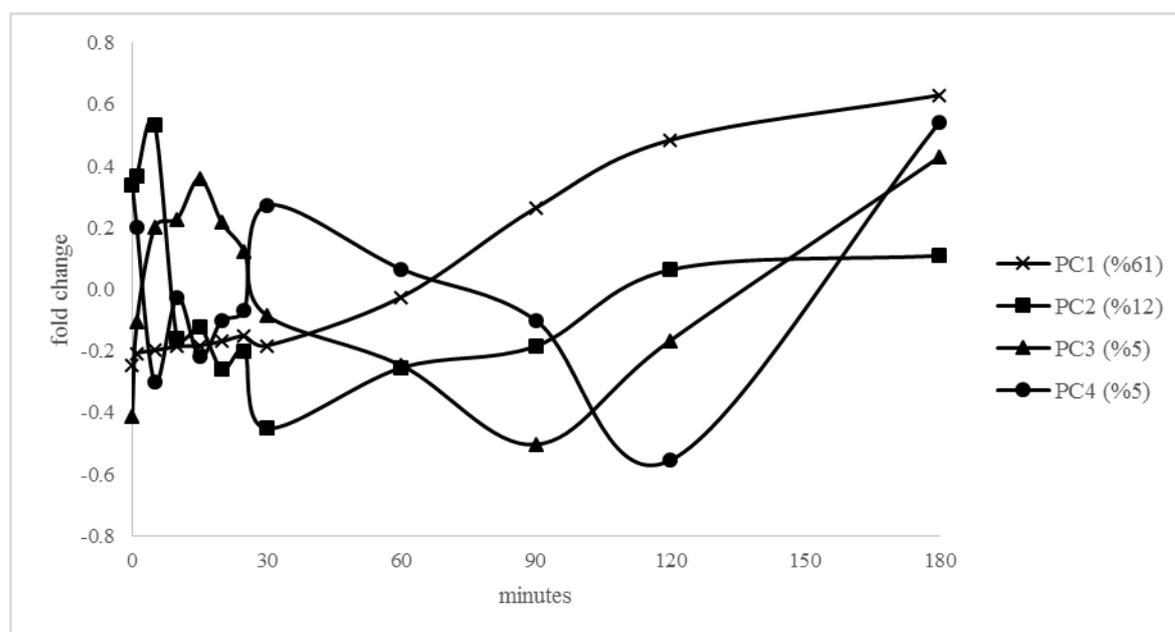


Figure 3.5. Splines of the expression profiles of the principal components identified by PCA. Percentages of the total variance explained by the principal components are written within the parenthesis.

The second principal (PC2) covers 12% of total variance and represents 499 genes out of which 264 genes are positively correlated with PC2 and were down-regulated within the first 30 minutes after the pulse. A partial recovery of the transcription levels of these genes was observed in the following 90 minutes. Genes whose expression positively tie in the

component engaged to stress responsive processes and the oppositely coordinated 235 genes are linked to growth related biological processes and pathways.

The third principal (PC3) represents 166 genes out of which 75 genes are positively correlated with PC and were up-regulated within the first 15 minutes, downregulated within the following 75 minutes and up-regulated again for the rest of the profile. Remaining 91 genes are negatively correlated with PC3.

The fourth principal (PC4) represents 39 genes of which 22 genes positively linked with the 4<sup>th</sup> principal whose expression profiles follow two successive valleys of down-regulation and subsequent up-regulation. The remaining 17 genes negatively correlated to PC4.

Principal component analysis is an efficient tool to extract largest group of genes with similar expression profiles from the transcriptome represented with PC1. However following principals are inclined to be hypothetical and loosely linked to real gene expression profiles. For instance the highest Pearson correlation coefficient between PC4 and any gene is only 0.834 implying that if 0.85 was chosen as similarity threshold then there would be no genes in the set linked to the principal.

3.3.2.3. Self-organizing maps (SOMs). In order to analyze the data in smaller groups SOMs were used. The number of desired clusters can be supervised in advance and the method is a well-established mathematical cluster analysis which has been used in microarray data analysis widely (Tamayo *et al.*, 1999).

A total of 1150 nodes with standard deviation more than 0.25 were used to generate SOMs. Time-course global transcriptome data collected after doxorubicin pulse was clustered into 36 clusters containing 6 to 72 genes (Figure 3.6). c0 cluster was found to be significantly associated with cellular amino acid biosynthetic process ( $p\text{-value}=5.0\text{E-}15$ ), translation ( $p\text{-value}=1.8\text{E-}4$ ) and oxidation reduction ( $p\text{-value}=8.5\text{E-}4$ ) processes and pathways related to amino acid metabolism. 5 clusters (c12, c13, c18, c19 and c24) were enriched in growth related terms like ribosome biogenesis ( $p\text{-values} \leq 1.16\text{E-}03$ ). Cluster c34 was identified to be associated with proliferation terms like sporulation ( $p\text{-value}=3.70\text{E-}$

03). Remaining 29 clusters have no significant enrichment with any GO term with a significance threshold of *0.001*.

In order to attain biological associations these clusters were merged. An analysis of the heat-map of the SOMs indicated that all clusters may be divided into two groups as up-regulating or down regulating genes with time.

In order to avoid over approximation a more stringent criteria was used for merging clusters. Cluster sets were merged whenever all possible pairwise correlations between cluster's centroids are strong enough ( $r > 0.90$ ) to form a new meta-cluster and the number of clusters that will remain unmerged was minimized.

A total of 34 clusters were merged into 6 meta-clusters leaving two outlier clusters; c28 and c29. Both of the outlier clusters have centroids indicating low expression profiles and both are not significantly associated with any biological role.

The largest meta-cluster, MC1 composed of 526 genes, was formed by merging of 15 clusters (c0-c5, c7-c11, c14, c16, c17 and c23). MC2 to MC6 consist of 8 clusters (c21, c27, c30-c35), 4 clusters (c18, c19, c24 and c25), 3 clusters (c6, c12 and c13), 2 clusters (c20 and c26) and 2 clusters (c15 and c22), respectively (Figure 3.6).

Meta-cluster, MC1 accounts for processes related to amino acid metabolism, transport and stress response. MC2 represents reproduction processes and meta-clusters from MC3, MC4 and MC5 significantly related to growth of the cell. Genes of the MC6 (c15 and c22) are not enriched in any specific biological process or pathway.

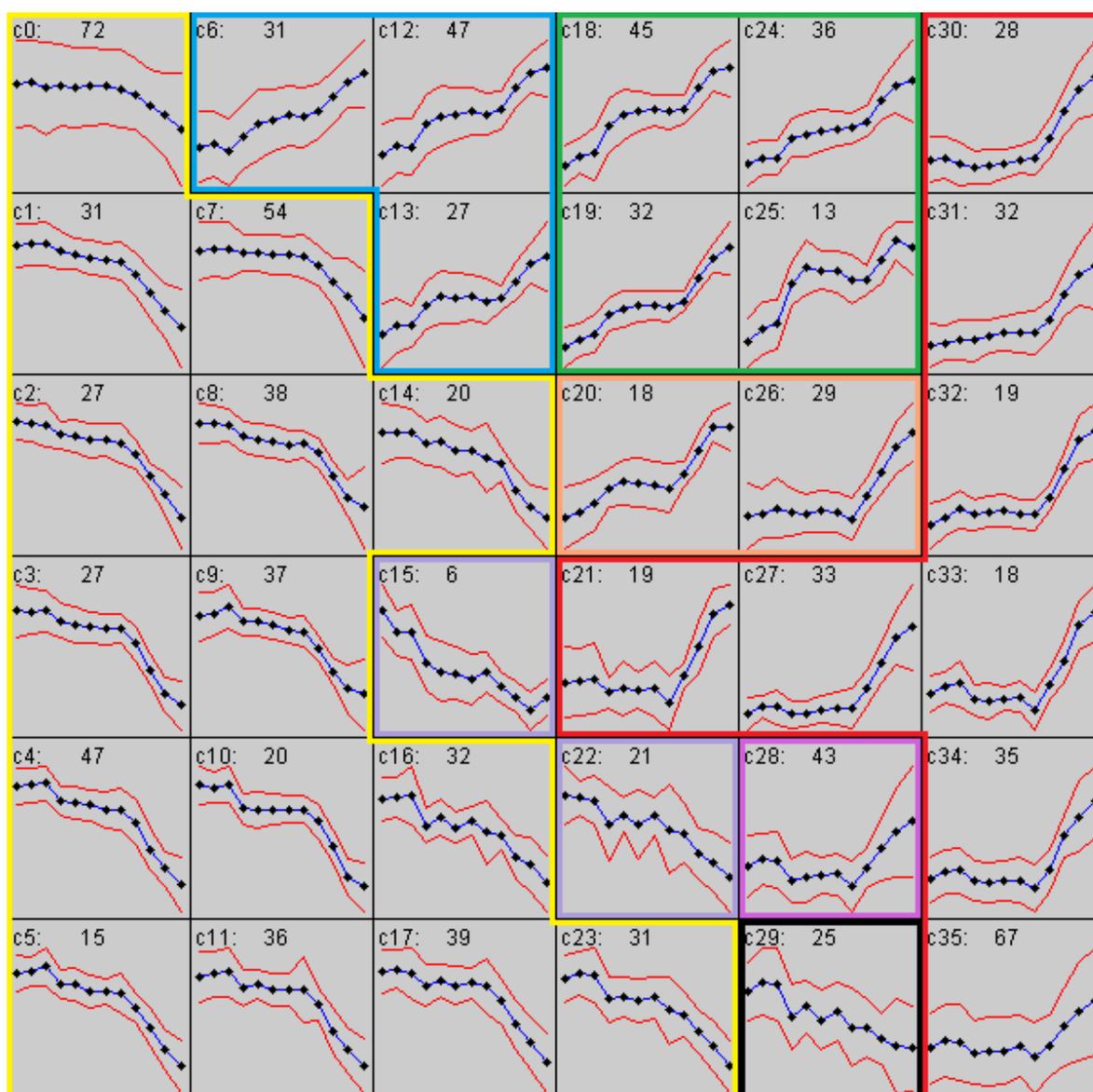


Figure 3.6. SOMs of transcriptome data representing dynamic response to doxorubicin stress. Numbers beginning with ‘c’ are cluster IDs and the second numbers are the number of nodes in the clusters. Meta-clusters MC1 to MC6 were outlined by colored borders.

3.3.2.4. Weighted Gene Co-expression Network Analysis (WGCNA). Weighted gene co-expression analysis (WGCNA) was found to one of the best approaches to construct global co-expression networks to analyze transcriptional profiles and for functional annotations (Zhang and Horvath 2005). Global transcriptome data was also analyzed by using WGCNA. Power function was selected as an adjacency function and parameterization of the power function was identified according to fitness of the resulting co-expression network to scale

free criteria. 6<sup>th</sup> power of the normalized Pearson correlation was assigned as weights of the pairwise interactions (Figure 3.7).

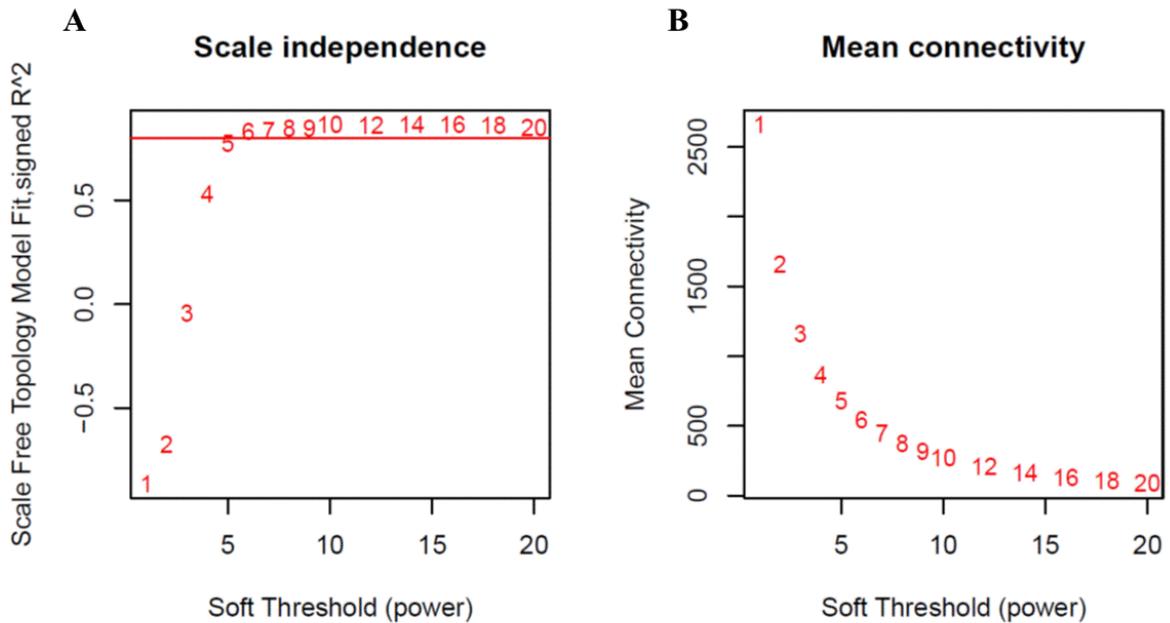


Figure 3.7. Fitness to scale-free nature (A) and mean connectivity (B) as a function of parameterization of the power function.

Resulting adjacency matrix transformed into a topological overlap matrix (TOM) and then dissimilarities based on TOM were hierarchically clustered by using WGCNA R Package. Minimum module size was set to be 30 and module pairs with higher correlation than 0.75 between their representative expression profiles were merged. 14 modules were identified representing the response of the cells to doxorubicin stress. 37 genes were identified as outliers; do not belong to any module (Figure 3.8). Each cluster was labeled by a color and module sizes were tabulated (Table 3.1).

Table 3.1. Module ID's and sizes determined by WGCNA.

<u>Module ID</u>	<u>Module Size</u>	<u>Module ID</u>	<u>Module Size</u>	<u>Module ID</u>	<u>Module Size</u>
turquoise	2472	red	176	green-yellow	75
blue	1008	black	174	tan	70
brown	861	pink	111	salmon	64
yellow	196	magenta	99	cyan	40
green	194	purple	78	grey (outliers)	37

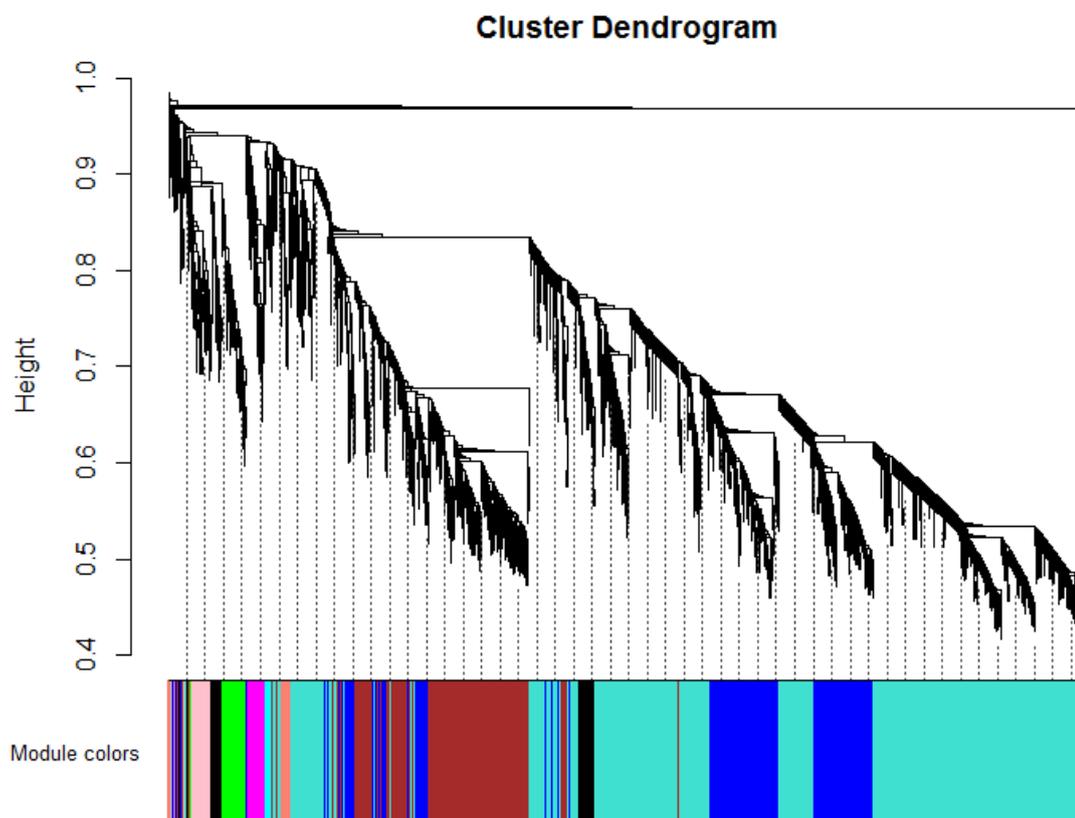


Figure 3.8. Hierarchical clustering dendrogram of 14 clusters by WGCNA.

Hierarchical network and heat map of the modules based on correlations between their eigengene profiles provide clues about the organization of the modules (Figure 3.9). The brown and the pink modules with short-term induction after doxorubicin addition to the media are clustered together. The green module displays opposite expression behavior compared to the brown module. Modules co-clustered with the green module like the tan module has no common ontological feature.

The genes within three modules; the yellow, the red and the magenta were down-regulated first after the pulse and they recover their original expression levels after 30 to 90 minutes. The red and the yellow modules belong to the same hierarchical branch. The magenta module was clustered with the cyan module. Cyan module consists of the genes which are not significantly associated with any biological process or pathway.

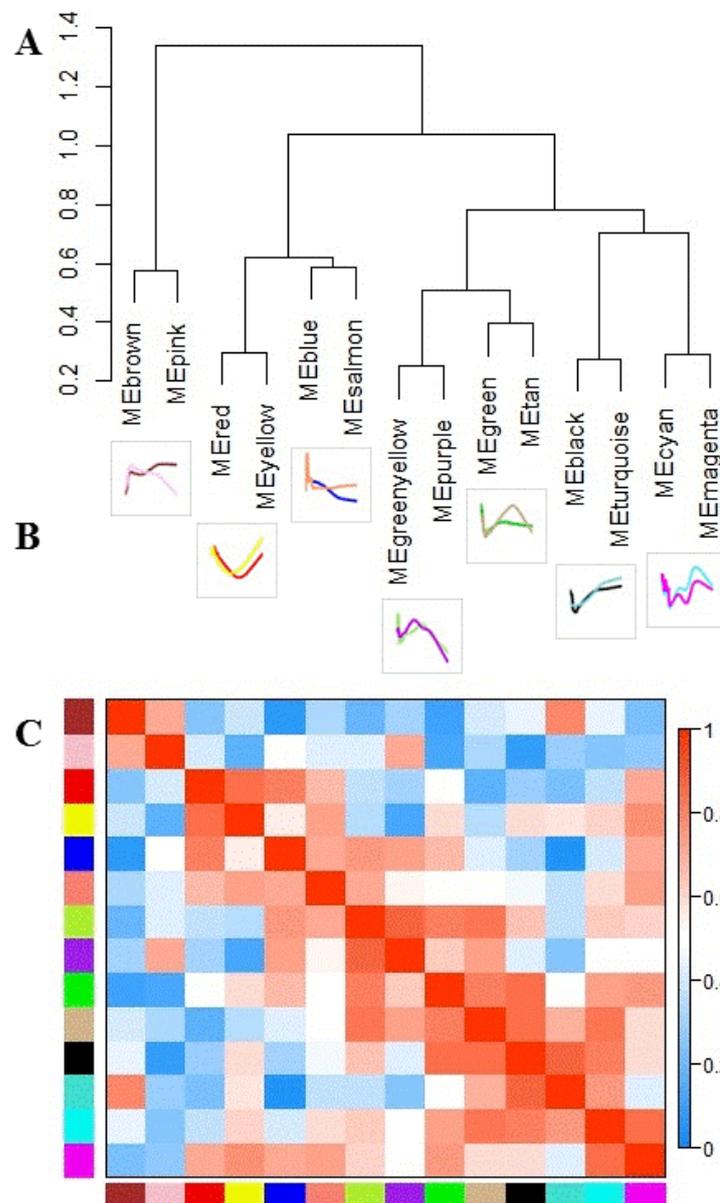


Figure 3.9. Meta-analysis of the modules identified by WGCNA. Hierarchical organization of the modules (A), splines of eigengene profiles (B) and heat-map of the modules (C).

The turquoise module, the largest module which includes almost half of the transcriptome, shows an up-regulation with a delay of 60 minutes. The second largest module is the blue module whose expression profile is just the opposite of the largest one. The purple module is clustered with the green module. The salmon module placed next to the blue module in the hierarchical tree consists mostly of genes displaying little variation in the level of expression and genes with very low level of expression and this module was not found to be associated with any biological process term. The closest neighbors of the

green-yellow module in the dendrogram the purple, the green and the tan modules. The genes of the tan module alone have no significant enrichment with any biological GO process term.

### **3.3.3. Integrative approaches**

It has been shown that integrating multi-omics data can unveil additional information about biological systems (de Keersmaecker *et. al.*, 2006). In order to analyze dynamic transcriptomic data of doxorubicin stress in yeast chemostat culture different integrative approaches were used to integrate transcriptomic response with interactome.

3.3.3.1. NP analysis. Integrating transcriptome with interactome is considered to be useful to extract both co-expressed and physically interacting group of genes/proteins. NP analysis maps transcriptome data onto protein-protein interaction network in order to identify large coordinated group of genes that mediate the dynamic response to environmental stresses (Xia *et. al.*, 2006).

Protein-protein interaction network of yeast consisting of 5487 proteins and 76838 directionless edges without loops was constructed and integrated with the dynamic transcriptomic data collected after the doxorubicin pulse into chemostat culture using NP analysis.

In NP analysis all edges should be labeled as positively correlated, negatively correlated or not correlated according to Pearson correlation coefficients (PCC) between expression profiles of the nodes connected by the edge. PCC values of 0.4 and -0.4 were described as cutoffs by the authors (Xia *et. al.*, 2006). PCC values of 0.7 and -0.7 were used as cutoffs in another study conducted in our laboratory (Dikicioglu *et. al.*, 2011). The effect of the selected values of PCC on the frequency of possible gene pairs in the transcriptome data and the effect of selected PCC threshold on variation of number of nodes and edges in NP network were analyzed in order to identify an acceptable threshold which can reduce the size of network to an amenable size (Figure 3.10).

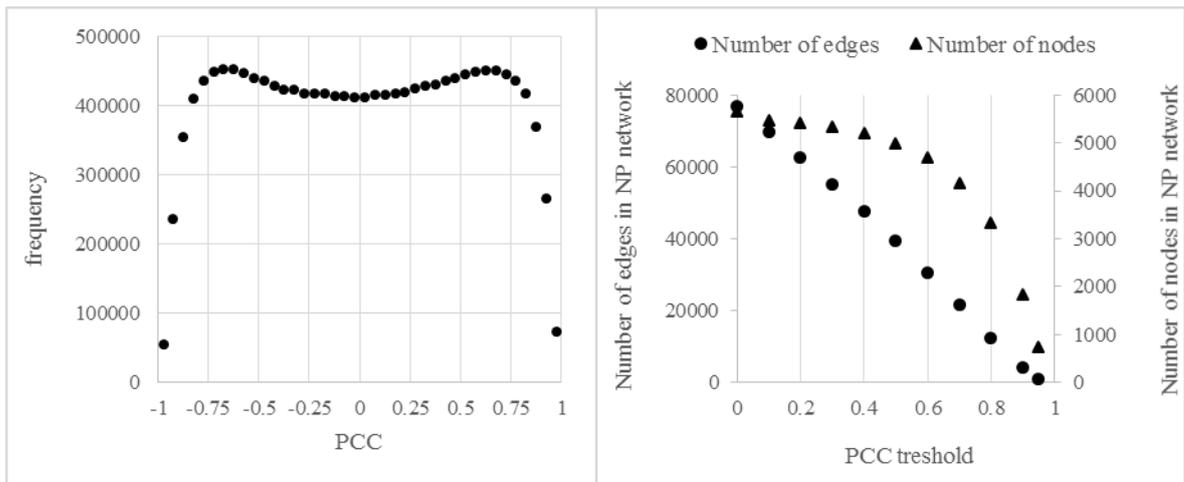


Figure 3.10. Frequency distribution of PCCs and network properties of NP network with respect to PCC threshold.

PCC cutoff was selected intuitively in the light of Figure 3.10 as 0.8 and -0.8 for correlation and anti-correlation respectively in NP analysis of the present data. Resulting NP network consists of 3333 nodes and 12472 edges. 8490 edges connect positively correlated and 3982 edges connect negatively correlated node couples in this network.

After hierarchical clustering of the expression profiles of the nodes present in NP network by HCE 3.0, modules with less than 1% intra-cluster anti-correlation edges were manually dissected. The network smoothly split into two modules without missing out any nodes and without violating the restriction defined for intra-cluster anti-correlation edges. Slightly larger module was named as R module with 1723 nodes and the other was named as B module (Table 3.2).

Table 3.2. Distribution of the nodes and anti-correlated edges in NP network.

	R module	B module
Number of nodes	1723	1610
Percent of anti-correlated interactions within cluster	0.21%	0.31%
Percent of correlated interactions between cluster	98.98%	

These modules were further split into two sub-parts; Interface modules (RI and BI) consisting of nodes which interact with the opposite module and core modules (RC and BC) consisting of nodes which have only intra-modular interactions. 105 nodes in NP network was disconnected from bulk of the network. Number of nodes in the modules were tabulated in Table 3.3.

Table 3.3. Number of nodes in 4 modules identified in NP.

Module ID	Number of nodes
RC	561
RI	1162
BI	1126
BC	484
Total (NP)	3333

Average expression profiles of the four modules identified by NP analysis were also plotted (Figure 3.11). The expression profiles are consistent with the expression profiles identified by hierarchical clustering of whole transcriptome data (Figure 3.3).

Hierarchical clustering of an active sub-network (NP network) instead of whole transcriptome provided additional information. Upregulation of genes involved in ribosome biogenesis and rRNA processing with time could be detected by NP analysis which integrates the transcriptome data with interactome. Analysis of global transcriptome data alone by hierarchical clustering could not reveal this term. Similarly the stress response related process terms identified in BC module, implying downregulation of the associated genes could be identified. The genes could not be detected by hierarchical clustering of the whole transcriptome.

### 3.3.4. Transcriptional Regulation

Biological systems regulate intracellular entities in a complex and dynamic way in response to changes in the media. Dynamic reorganization of the transcriptome in response to a doxorubicin pulse was investigated using DREM in order to identify key TFs involved in the response in time dependent manner.

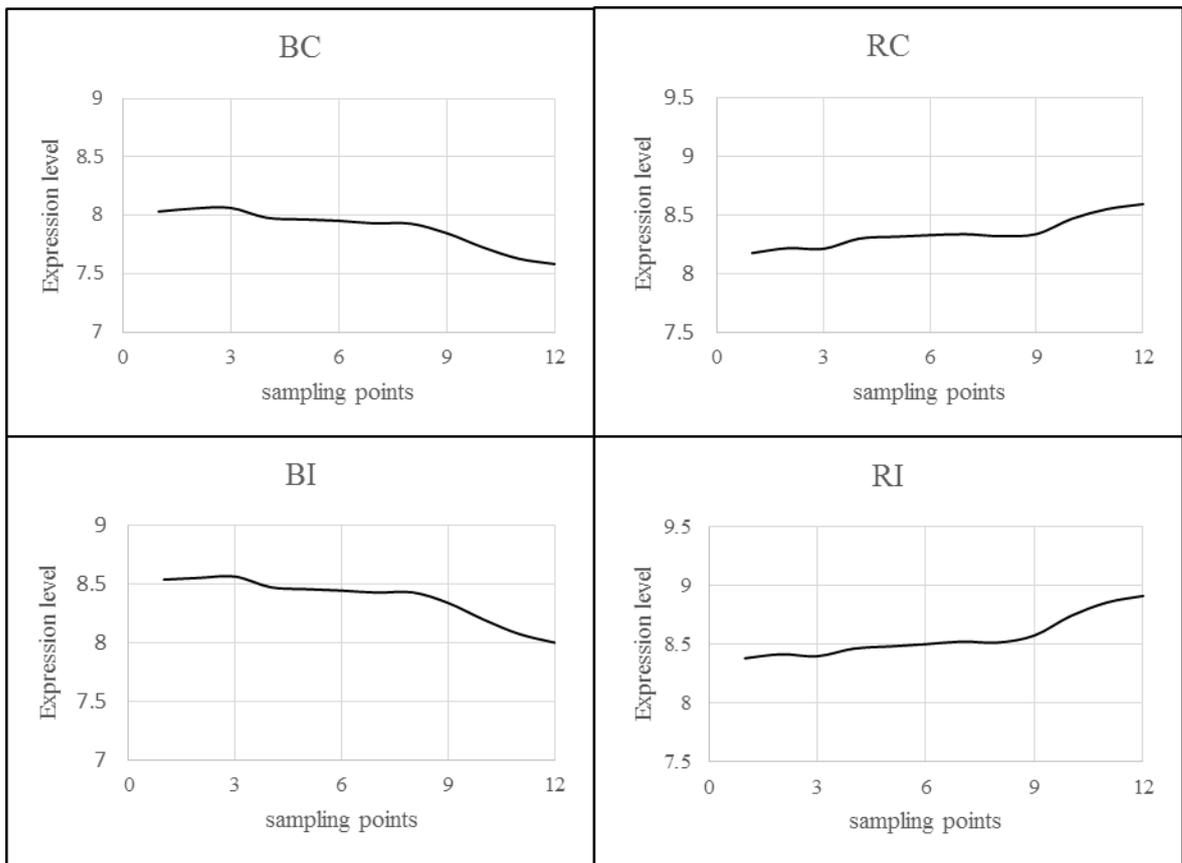


Figure 3.11. Average expression profiles of the modules identified by NP analysis.

Dynamic Regulatory Event Miner (DREM) was used in order to infer an annotated global temporal map of the cellular dynamic regulation. This method integrates transcriptome data with DNA-protein interaction network. In this study a network of 7840 regulatory links between 118 transcriptional factors (TFs) and 5971 genes (MacIsaac *et al.*, 2006) was used.

Expression profiles of 2299 genes out of whole transcriptome with known TF interactions were identified to have at least 1.5 fold change between their transcriptional maxima and minima. Analysis of this subset by DREM revealed that doxorubicin pulse triggers a complex regulatory mechanism where 11 transcription factors (TFs) drives transcriptome into 9 branches (Figure 3.12).

Immediate split in the transcriptional response was found to be mediated significantly by regulators of cell cycle progression (Mbp1p) and amino acid biosynthesis (Gcn4p) (Figure 4.12). After a second split occurring between 1<sup>st</sup> and 5<sup>th</sup> minutes, the most prominent

bifurcation took place between 5<sup>th</sup> and 10<sup>th</sup> minutes which is mediated by TFs regulating DNA replication and repair (Abf1p) and oxidative stress response (Skn7p) (Figure 3.12 - 3.13).

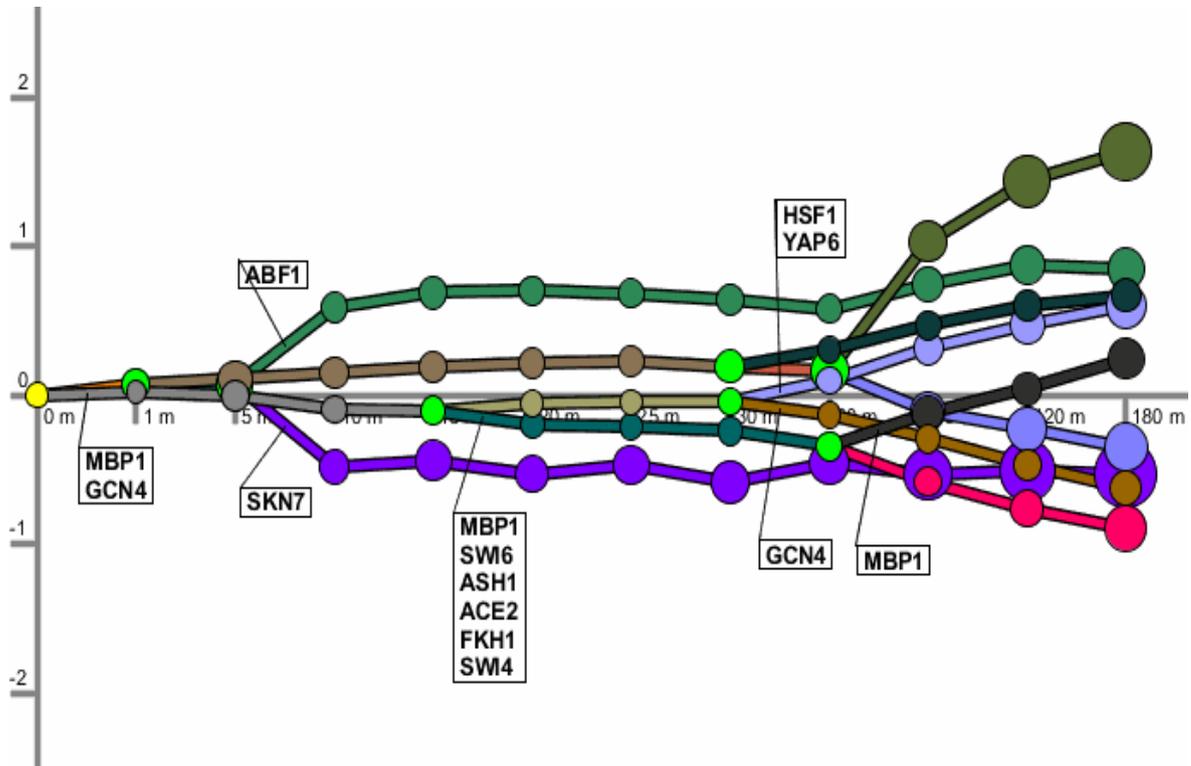


Figure 3.12. Dynamic regulatory map of yeast cells in response to doxorubicin pulse. X-axis was arranged as uniform sample points instead of real time scale in order to prevent visual ambiguity. Light-green dots indicate the location of bifurcations.

Green and violet branches visualized in Figure 3.13 were associated with 226 and 242 gene expression profiles, respectively. Genes in the upregulated branch are strongly enriched in growth related GO terms like ribosome biogenesis ( $p\text{-value}=1.4E-105$ ). Downregulated profiles were found to be related to stress response. These two extreme profiles do not show any split afterwards.

After 15 minutes the grey branch undergoes a minor split under influence of transcriptional factors related to cell cycle (Mbp1p, Swi4p, Swi6p, Ash1p, Ace2p and Fkh1p) (Figure 3.12). Resulting in upper-child and lower-child branches which bifurcates again after 30 and 60 minutes, respectively (Figure 3.14-15).

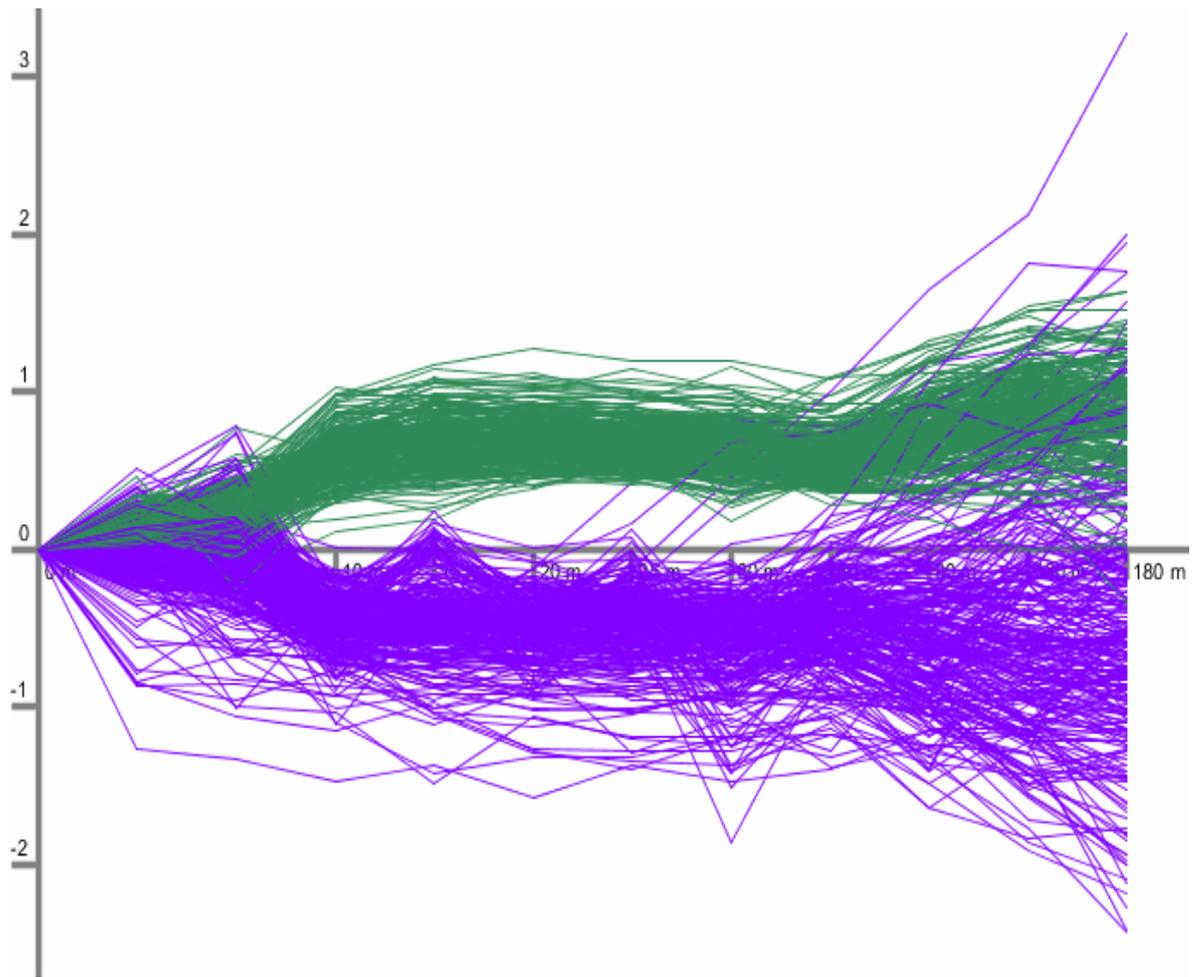


Figure 3.13. Bifurcation of green and violet branches 5 minutes after doxorubicin addition

322 genes of Lilac branch which upregulated after 30 minutes by heat shock transcriptional factor (Hsf1p) and regulator of transcription from RNA polymerase II (Yap6p) was found to be significantly associated with protein folding ( $p\text{-value}=4.7E-6$ ) and regulation of protein catabolic process ( $p\text{-value}=8.0E-4$ ). Brown downward branch of 429 genes regulated by activator of amino acid biosynthesis (Gcn4p) was identified to be enriched in cellular amino acid biosynthetic process ( $p\text{-value}=1.9E-5$ ) (Figure 3.14).

Dark green branch splits at 60<sup>th</sup> minute under the control of Mbp1p which is involved in the regulation of cell cycle progression, 133 genes of black sub-branch was identified to be associated with DNA replication and repair ( $p\text{-value}=5.6E-11$ ;  $p\text{-value}=3.7E-10$ ). The red branch of 341 profiles was enriched in GO biological process terms like translational elongation ( $p\text{-value}=3.8E-5$ ) and cellular response to stress ( $p\text{-value}=5.2E-4$ ) diverges downwards after 60 minutes (Figure 3.15).

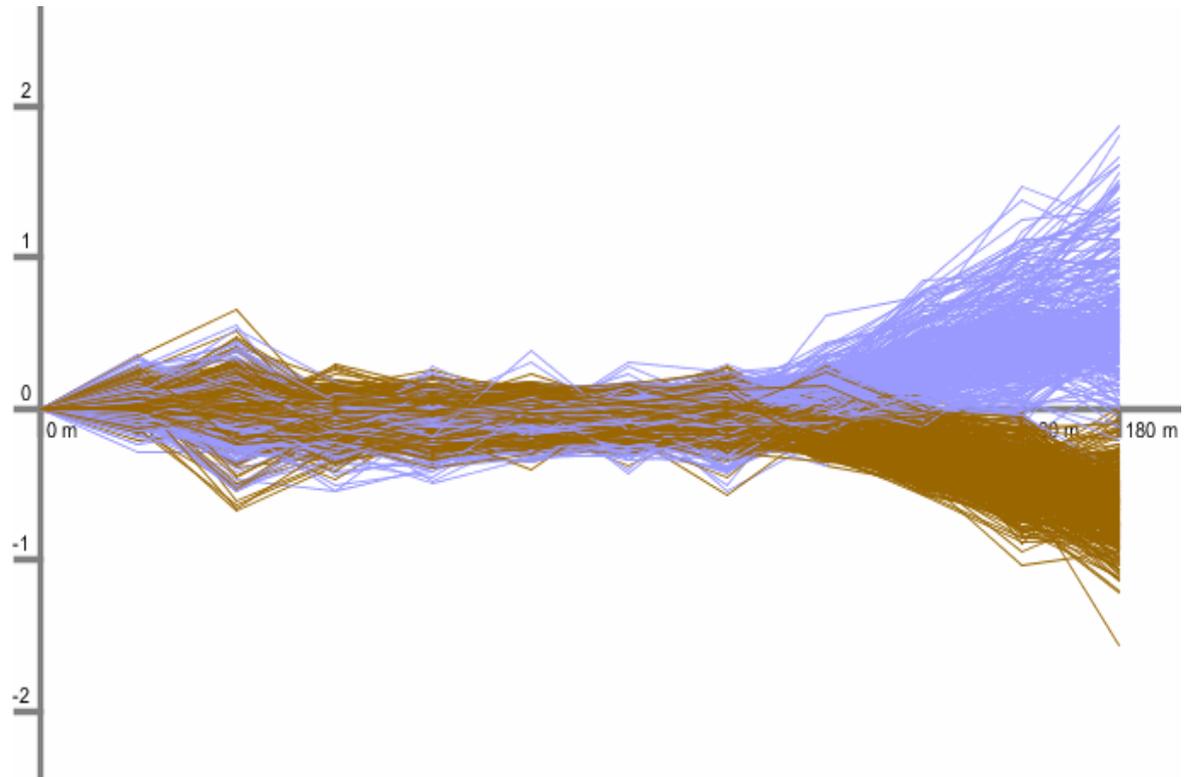


Figure 3.14. Bifurcation of the khaki branch 30 minutes after doxorubicin addition.

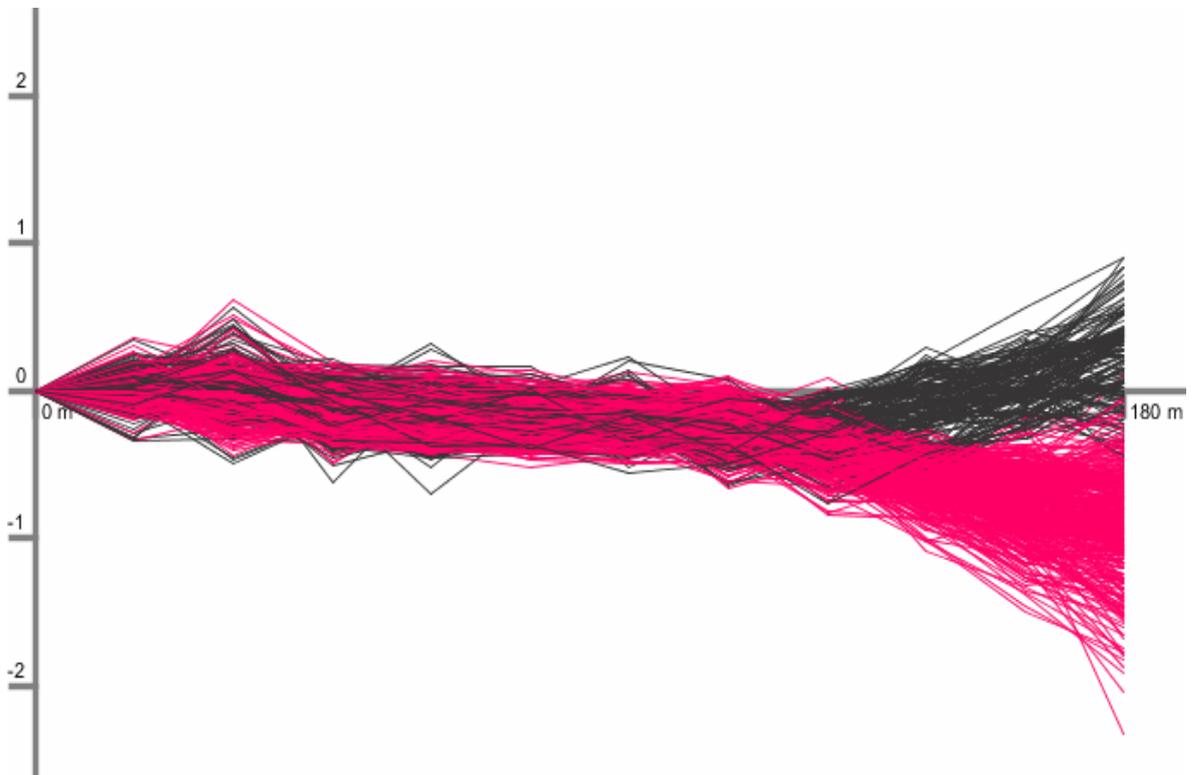


Figure 3.15. Bifurcation of the dark green branch 60 minutes after doxorubicin addition.

The last branch separated from the rest after 1 minute splits into two branches at 30 minutes; petrol blue branch with 329 genes progress upwards and the other branch undertake another bifurcation at 60 minutes inducing olive green branch displaying the highest fold change at the end of the experiment and reducing blue branch (Figure 3.16).

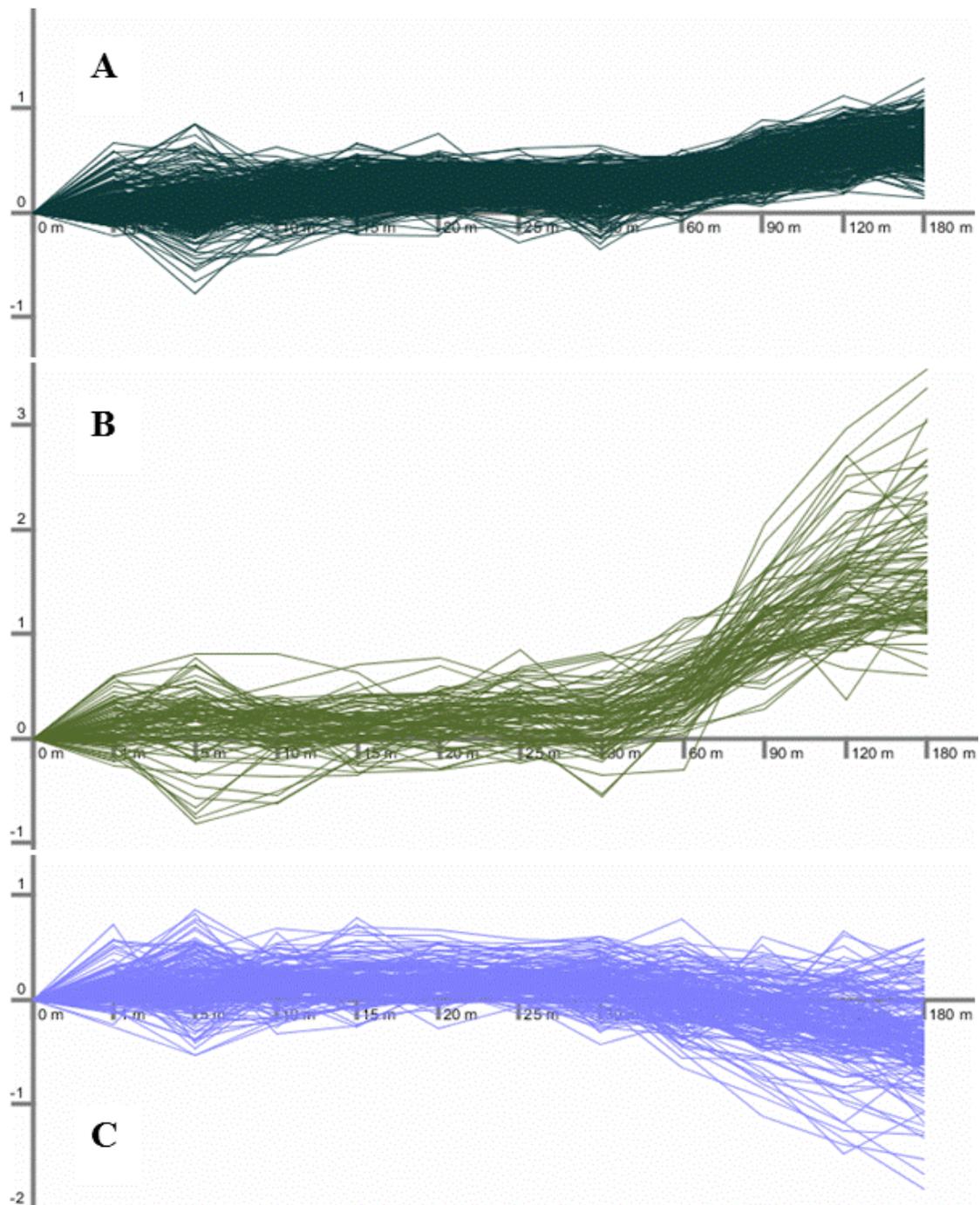


Figure 3.16. Expression profiles of Petrol Blue, Olive Green and Blue branches.

Petrol Blue, Olive Green and Blue branches were found to be significantly associated with GO biological process terms ncRNA processing ( $p\text{-value}=3.4E-12$ ), spore wall assembly ( $p\text{-value}=2.2E-11$ ) and purine nucleoside monophosphate biosynthetic process ( $p\text{-value}=2.8E-6$ ), respectively. Off-the-chart upregulated Olive Green branch and Blue branch consist of 91 and 186 genes (Figure 3.16).

### 3.3.5. Differentially expressed genes

Biological systems alter expression levels of their genes between varying biological conditions. It is important to identify differentially expressed genes. One of the most common method of significance analysis of time course datasets is EDGE approach. In this study differentially expressed genes were identified by EDGE and the approach was extended in order to enlarge the set of significant genes. Resulting significant subset were reanalyzed by SOMs, WGCNA, DREM and NP analysis.

3.3.5.1. Identification of Differentially and significantly expressed genes by EDGE. In order to identify genes whose expressions were significantly altered in a dynamic manner, EDGE (Extraction and analysis of differential gene expression) software package which was specifically designed for the analysis of time series datasets, was used (Leek *et al.*, 2006).

A total of 1658 genes was found to be differentially and significantly expressed after a doxorubicin pulse injected into the yeast chemostat culture ( $q\text{-value}<1.0E-4$ ) by using R package of EDGE. The analysis of these genes revealed that 765 genes were up- and 893 genes were down- regulated in response to the pulse. The up-regulated genes were found to be significantly associated with protein localization into organelle, protein import, siderophore-iron transport, iron assimilation by chelation and transport, iron transport, ncRNA processing, tRNA modification, translation and protein folding biological process terms. Proteasome pathway was observed to be enriched with this group of genes (Table 3.4).

The genes significantly associated with transcription and regulation of transcription, cellular amino acid catabolic process regulation of nitrogen compound metabolic process, cell wall organization, glycine, lysine, arginine biosynthetic processes, aging, intracellular

signaling cascade, response to unfolded protein and drug transport biological process GO terms were down regulated (Table 3.4).

Table 3.4. Selected GO biological process terms and pathways significantly associated with differentially expressed genes identified by EDGE.

<b>up- regulated genes</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
siderophore transport	8	1.6E-05
tRNA modification	20	6.4E-04
ncRNA processing	62	2.2E-03
iron assimilation by chelation and transport	5	3.3E-03
protein targeting to membrane	16	4.4E-03
protein import into nucleus	16	4.4E-03
translation	110	4.4E-03
intracellular protein transmembrane transport	22	4.7E-03
iron ion transport	11	6.6E-03
rRNA catabolic process	9	7.9E-03
protein folding	24	7.9E-03
modification-dependent macromolecule catabolic process	47	9.6E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	19	1.4E-07
<b>Down-regulated genes</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
positive regulation of transcription from RNA polymerase II promoter	32	5.2E-05
cellular amino acid catabolic process	17	2.8E-04
amine catabolic process	18	2.9E-04
cell wall organization	60	7.6E-04
amine biosynthetic process	38	1.1E-03
chromatin assembly or disassembly	21	1.5E-03
negative regulation of transcription from RNA polymerase II promoter	23	2.6E-03
regulation of cell cycle process	28	2.8E-03
translational elongation	74	3.3E-03
lysine biosynthetic process via aminoadipic acid	6	3.6E-03
nucleosome organization	19	4.1E-03
aging	22	4.3E-03
intracellular signaling cascade	41	6.3E-03
arginine metabolic process	8	7.5E-03
response to unfolded protein	12	7.8E-03
drug transport	9	8.2E-03
cell aging	19	8.7E-03

### 3.3.6. Development of an extended approach to identify differentially and significantly expressed genes

In order to include the differentially expressed genes which might be missed by EDGE into further analysis 1658 genes which were identified as differentially and significantly expressed by EDGE, 2299 genes which have at least 1.5 fold change between their maxima and minima used in DREM analysis and 1150 genes which have standard deviation of 0.25 or higher used in SOM analysis were comparatively investigated. This analysis revealed that 628 genes were commonly identified by these three approaches and a total of 2856 genes were identified (Figure 3.17).

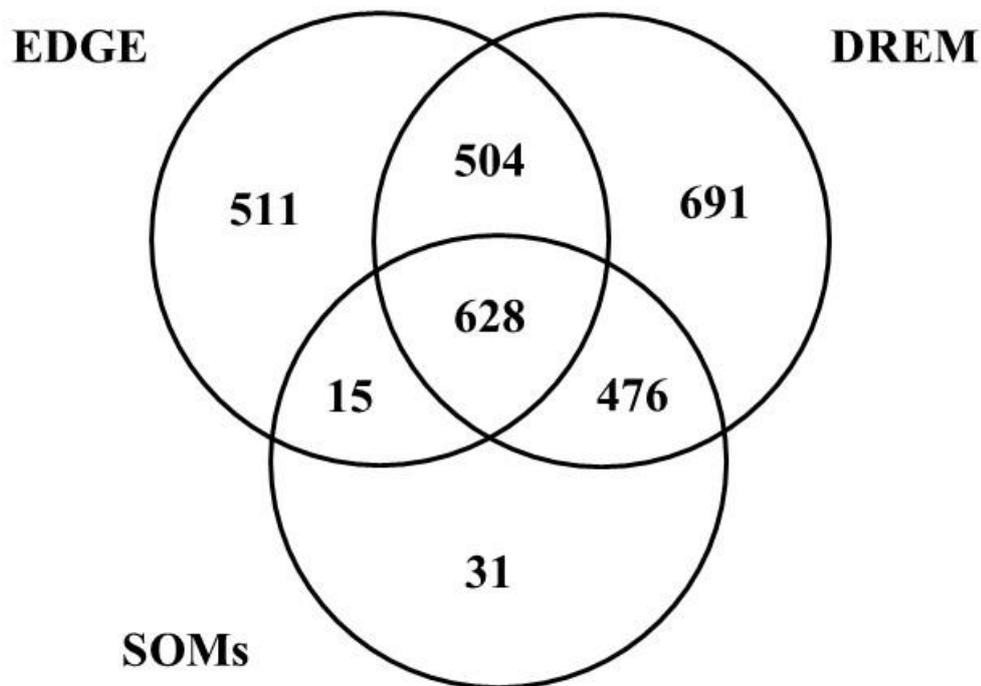


Figure 3.17. Venn diagram of significant genes identified by EDGE and used in DREM and SOMs analysis.

Microarray datasets can carry considerable noise and it has been shown that this noise is inversely proportional to RNA abundance (Zeisel *et al.*, 2010). In order to eliminate these false positives frequency distribution of average expression levels of the genes from conjugated three clusters were investigated (Figure 3.18).

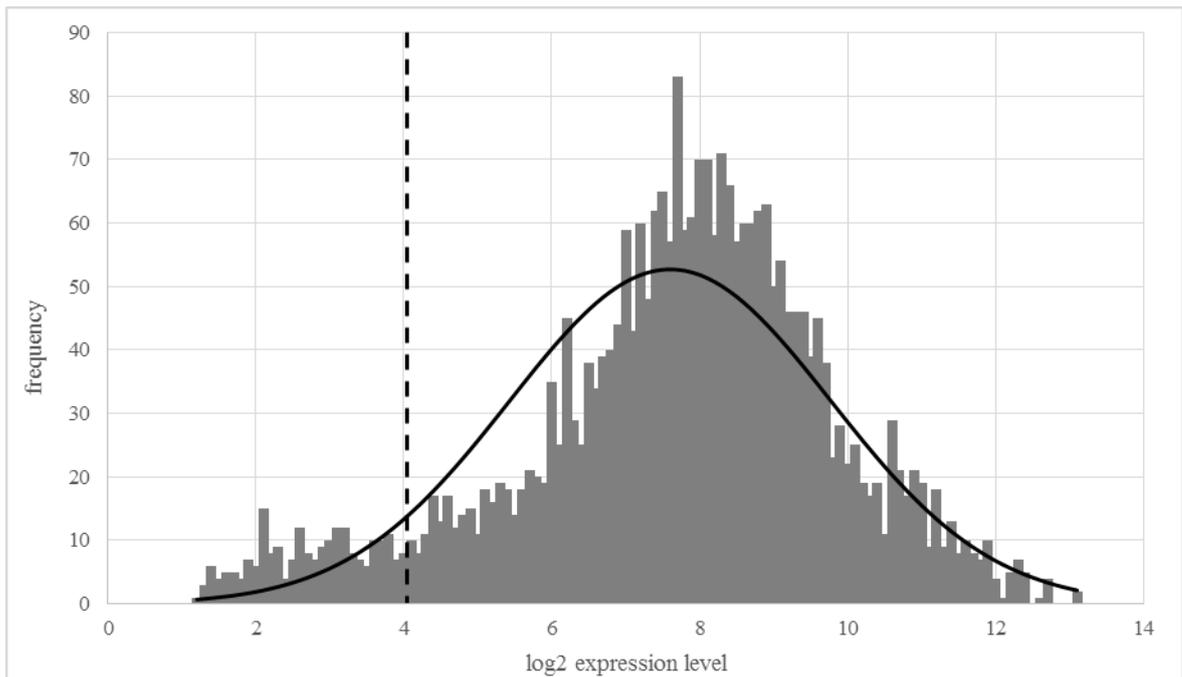


Figure 3.18. Frequency distribution of log2 ratios of average expression levels of the 2856 genes resides in at least one of three sets identified in EDGE, SOMs and DREM analysis.

The curve represents normal distribution and dashed line marks p-value = 0.05.

The frequency distribution of average expression levels of the 2856 genes was investigated and by fitting a normal distribution a threshold of 4.05 for the minimum average expression level which corresponds to the typical p-value of 0.05 was selected. A total of 226 out of 2856 genes (7.9 %) were then filtered in order to remove noise. 9.2 % and 11.3 % of the genes from the subsets originated from DREM and SOM analysis respectively, were removed to decrease the number of false positives. 2.4 % of the genes identified by EDGE was removed similarly. These results indicated that the subsets used in SOMs and DREM analysis carry more false positives.

Remaining 2630 genes were identified as differentially expressed and this extended EDGE (E-EDGE) approach was used for further analysis.

3.3.6.1. SOMs of differentially expressed genes identified by E-EDGE. In section 4.3.2.3 36 clusters identified by SOMs were investigated and then merged to have six meta-clusters for a subset of 1150 genes. In the present section manual merging was not applied and number of clusters was minimized to slightly higher than the number of meta-clusters

identified in section 4.3.2.3. The number of clusters was determined as nine, intuitively for SOMs of 2630 genes which were identified by E-EDGE.

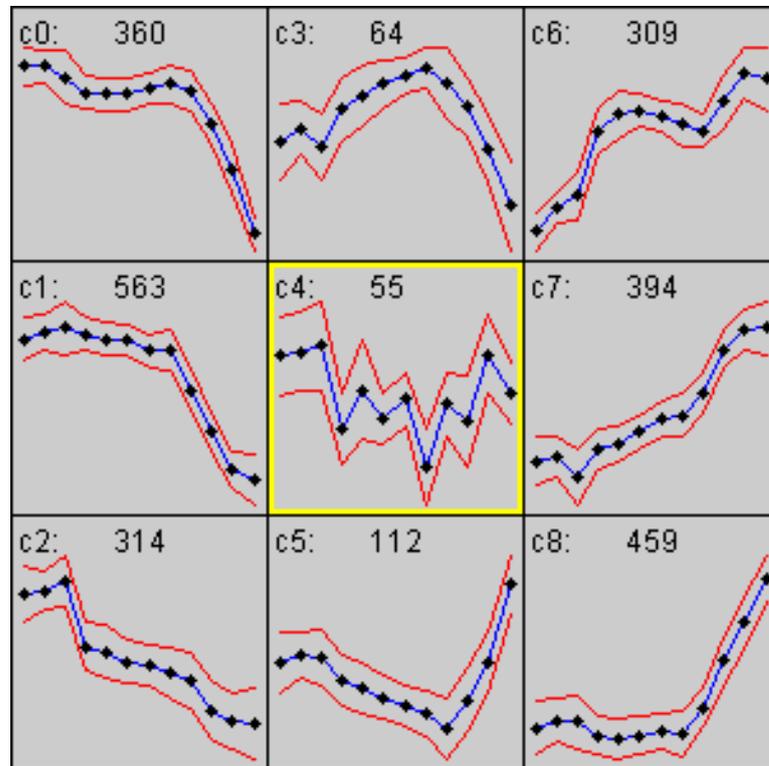


Figure 3.19. Self-organizing maps of time series transcriptome data representing dynamic response of differentially expressed genes identified by E-EDGE.

c0 and c1 clusters display a down-regulation with a delay of 60 and 30 minutes respectively after the pulse. Down-regulation of c2 module takes place 5 minutes after the pulse (Figure 3.19). The significantly associated biological process terms and pathways are presented in Table 3.5.

The smallest two clusters c3 and c4 have 64 and 55 genes respectively. Genes clustered in c3 experienced an induction in mid-term and a recovery in long-term and genes clustered in c4 displays the opposite behavior with the highest noise (Figure 3.19). The significantly associated biological process terms and pathways are presented in Table 3.6.

Table 3.5. Selected GO biological process terms and pathways significantly associated with downregulated clusters; c0, c1 and c2. The number of the genes was given in parenthesis.

<b>c0 (360)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
amine biosynthetic process	25	2.5E-06
arginine biosynthetic process	7	7.4E-06
ornithine metabolic process	5	3.8E-04
histidine biosynthetic process	5	2.9E-03
glycerophospholipid metabolic process	12	6.8E-03
water-soluble vitamin biosynthetic process	10	8.6E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Histidine metabolism	6	8.6E-04
Arginine and proline metabolism	7	3.9E-03
Pyruvate metabolism	7	7.6E-03
<b>c1 (563)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
cell wall organization	49	9.9E-06
regulation of cell cycle	34	4.6E-04
'de novo' IMP biosynthetic process	6	9.8E-04
amine catabolic process	13	1.4E-03
negative regulation of transcription from RNA polymerase II promoter	18	1.4E-03
positive regulation of transcription from RNA polymerase II promoter	21	1.6E-03
cellular amino acid catabolic process	12	2.1E-03
regulation of cell size	21	5.3E-03
aging	16	7.0E-03
intracellular signaling cascade	29	8.4E-03
response to osmotic stress	20	8.4E-03
chromosome organization	55	9.8E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Meiosis	24	7.70E-03
<b>c2 (314)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
vacuolar protein catabolic process	34	8.5E-17
cellular response to heat	41	5.1E-16
glycoside biosynthetic process	6	6.5E-06
trehalose biosynthetic process	6	6.5E-06
cellular carbohydrate catabolic process	14	1.0E-04
energy reserve metabolic process	10	2.1E-04
response to osmotic stress	15	7.4E-04
alcohol catabolic process	11	1.5E-03
glucose metabolic process	13	4.6E-03
regulation of protein kinase activity	7	6.8E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Starch and sucrose metabolism	14	1.2E-08

Table 3.6. Selected GO biological process terms and pathways significantly associated with the smallest clusters c3 and c4. The number of the genes was given in parenthesis.

<b>c3 (64)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
sulfur amino acid biosynthetic process	5	6.8E-04
methionine metabolic process	5	9.2E-04
biotin biosynthetic process	3	1.8E-03
pyridoxine biosynthetic process	3	5.1E-03
tRNA modification	5	6.5E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Selenoamino acid metabolism	4	1.9E-03
Cysteine and methionine metabolism	4	6.7E-03
<b>c4 (55)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
histone modification	5	6.7E-03

112 genes which compose cluster c5 display a sharp upregulation at 60<sup>th</sup> minute after a slight downregulation observed in mid-term. Cluster c8 which consists of 459 genes has also a striking upregulation in long-term following a half hour of steady expression after doxorubicin impulse. The biological process GO terms and pathways associated with these two cluster was given in Table 3.7.

Table 3.7. Selected GO biological process terms and pathways significantly associated with clusters c5 and c8 which shows an upregulation in long term. The number of the genes was given in parenthesis.

<b>c5 (112)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
DNA repair	23	1.5E-09
mitotic sister chromatid cohesion	10	3.4E-09
M phase of mitotic cell cycle	15	8.9E-06
lagging strand elongation	6	3.5E-05
M phase of meiotic cell cycle	14	6.0E-05
regulation of DNA replication	6	3.7E-04
response to abiotic stimulus	17	1.1E-03
reciprocal meiotic recombination	6	2.5E-03
cell division	16	3.3E-03
regulation of gene expression, epigenetic	8	5.1E-03
gene silencing	8	5.4E-03
premeiotic DNA synthesis	3	5.6E-03
spindle pole body organization	4	6.8E-03
regulation of catabolic process	6	7.0E-03
reproductive process in single-celled organism	10	7.9E-03

Table 3.7. Selected GO biological process terms and pathways significantly associated with clusters c5 and c8 which shows an upregulation in long term. The number of the genes was given in parenthesis. Cont.

regulation of cell cycle	10	8.5E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
DNA replication	7	8.4E-05
Mismatch repair	6	8.7E-05
Proteasome	7	2.1E-04
Cell cycle	12	2.2E-04
Base excision repair	4	9.0E-03
<b>c8 (459)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
regulation of protein catabolic process	8	1.3E-03
siderophore transport	5	2.7E-03
ascospore formation	18	3.5E-03
protein refolding	6	6.3E-03
ubiquitin-dependent protein catabolic process	24	8.9E-03
response to pheromone	16	9.7E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	13	1.4E-06

Upregulation of 309 genes which come into c6 starts immediately after the pulse while upregulation of cluster c7 which consists of 394 genes begins with 5 minute delay. The biological process GO terms and pathways associated with these two cluster was given in Table 3.8.

Table 3.8. Selected GO biological process terms and pathways significantly associated with the upregulated clusters c6 and c7. The number of the genes was given in parenthesis.

<b>c6 (309)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ribosome biogenesis	161	2.5E-121
ribosomal large subunit assembly	26	1.0E-22
rRNA modification	10	5.0E-08
regulation of translation	31	8.1E-08
tRNA methylation	9	2.9E-06
ribosomal large subunit export from nucleus	8	1.4E-05
positive regulation of transcription from RNA polymerase I promoter	7	2.4E-05
snRNA pseudouridine synthesis	5	1.5E-04
snoRNA 3'-end processing	7	1.2E-03
peptidyl-diphthamide biosynthetic process from peptidyl-histidine	4	1.7E-03
one-carbon metabolic process	13	2.1E-03
exonucleolytic trimming during rRNA processing	6	3.7E-03

Table 3.8. Selected GO biological process terms and pathways significantly associated with the upregulated clusters c6 and c7. The number of the genes was given in parenthesis.

Cont.

U4 snRNA 3'-end processing	5	5.1E-03
RNA localization	16	7.9E-03
translational initiation	8	9.3E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
RNA polymerase	14	4.7E-12
Pyrimidine metabolism	19	9.9E-12
Purine metabolism	17	6.7E-08
Ribosome	18	1.0E-06
<b>c7 (394)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
methionine biosynthetic process	11	3.0E-05
protein import into nucleus	14	8.0E-05
sulfate assimilation	6	2.6E-04
amine biosynthetic process	21	1.3E-03
coenzyme biosynthetic process	13	2.3E-03
ncRNA 3'-end processing	9	2.3E-03
nonfunctional rRNA decay	6	3.2E-03
tRNA catabolic process	6	3.2E-03
cellular iron ion homeostasis	9	3.3E-03
RNA transport	17	3.3E-03
cellular metabolic compound salvage	12	3.4E-03
protein targeting to membrane	11	4.4E-03
regulation of translation	23	4.6E-03
polyadenylation-dependent ncRNA catabolic process	6	5.7E-03
snoRNA processing	8	5.7E-03
cysteine biosynthetic process	5	6.6E-03
tRNA wobble uridine modification	7	6.7E-03

**3.3.6.2. WGCNA of differentially expressed genes identified by E-EDGE.** A co-expression network was constructed using the subset of global transcriptome data identified by E-EDGE and analyzed by WGCNA.

Power function was selected as an adjacency function and parameterization of the power function was identified according to fitness of the resulting co-expression network to scale free criteria. Scale free topology could not be attained for reasonably low values of  $\beta$  and it was set simply as 6 following the recommendation of the authors of the approach as described in the user manual of WGCNA (Langfelder and Horvath 2008) (Figure 3.20).

Resulting adjacency matrix transformed into a topological overlap matrix (TOM) and then dissimilarities based on TOM were hierarchically clustered by using WGCNA R Package. Minimum module size was set to be 10 and module pairs with higher correlation than 0.95 between their representative expression profiles were merged.

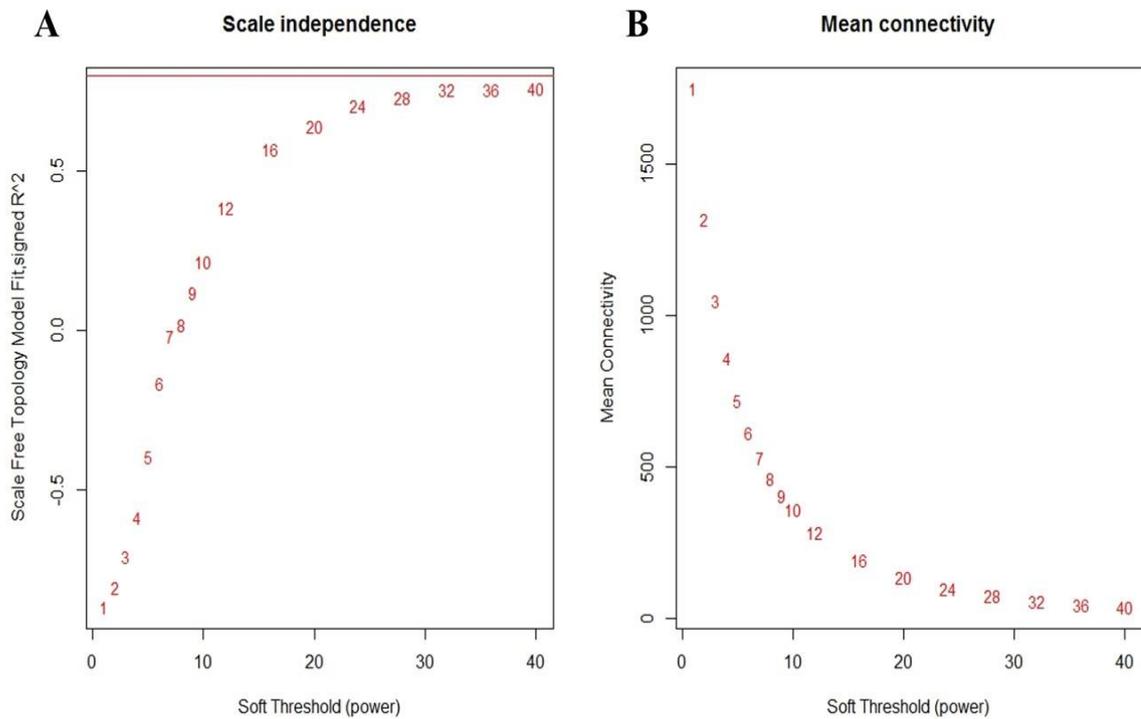


Figure 3.20. Fitness to scale-free nature (A) and mean connectivity (B) as a function of parameterization of the power function.

A total of 12 modules were identified representing the response of the cells to doxorubicin stress. No gene was identified as outlier; all genes belong to a module (Figure 3.21). Each cluster was named by a color and module sizes were tabulated (Table 3.9).

Table 3.9. Module ID's and sizes of E-EDGE genes determined by WGCNA.

<u>Module Name</u>	<u>Module Size</u>	<u>Module Name</u>	<u>Module Size</u>	<u>Module Name</u>	<u>Module Size</u>
turquoise	904	green	105	magenta	20
blue	594	red	66	purple	19
brown	417	black	45	green-yellow	17
yellow	389	pink	38	tan	16

Hierarchical network and heat map of the modules based on correlations between their eigengene profiles provide clues about the organization of the modules. Modules are organized in three groups. The first group consists of green-yellow, brown and tan modules, the second group which has the closest correlation between their eigengenes consists of blue, red, black and yellow modules and the last group which has the weakest correlation among eigengene profiles of its members consists of pink, purple, magenta, green and turquoise modules (Figure 3.22).

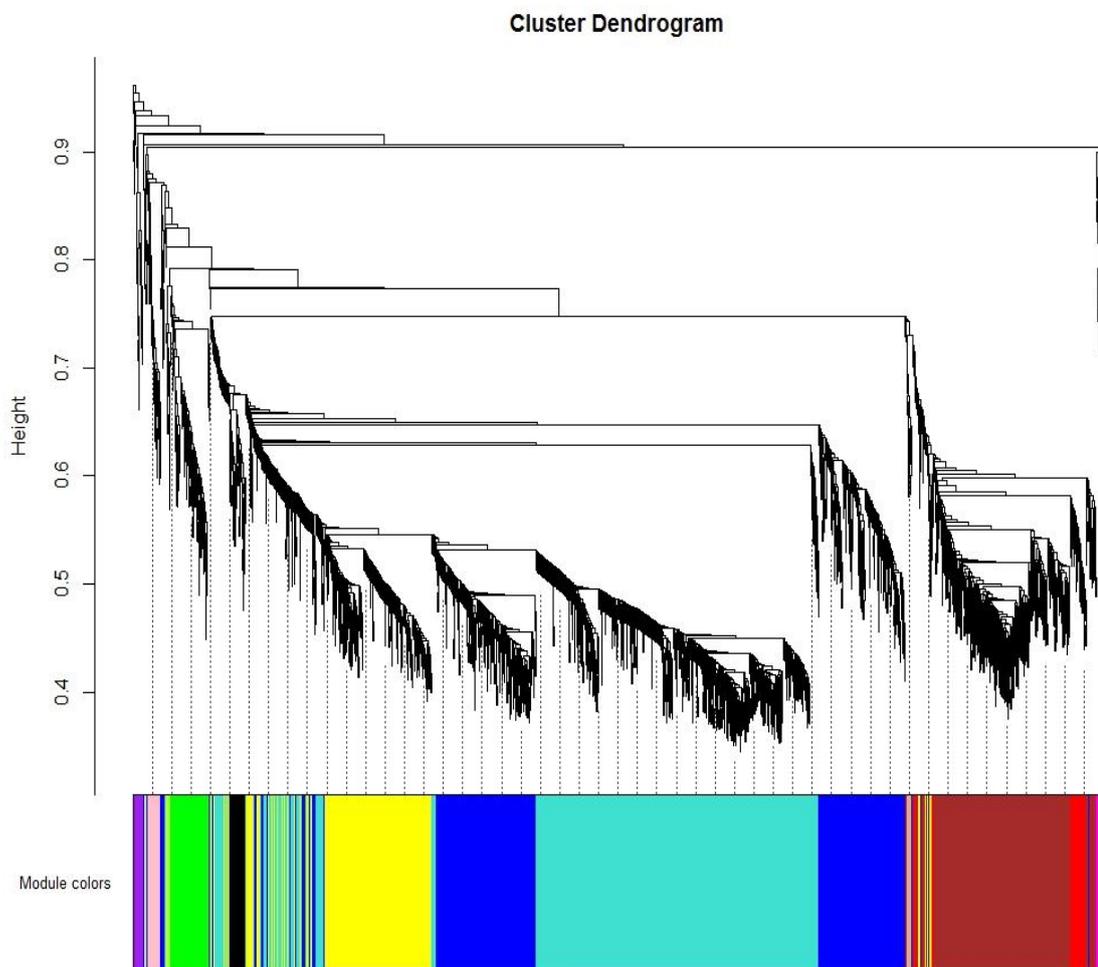


Figure 3.21. Hierarchical clustering dendrogram of 12 clusters of E-EDGE genes by WGCNA.

Eigengene profiles were also plotted in order to visualize similarities between modules to get general picture about the dynamic response of different modules identified by WGCNA. All of the first group of modules (green-yellow, brown and tan) identified in

Figure 3.22 has a change in their expression levels 5<sup>th</sup> minutes after the doxorubicin pulse. The main expression change in the second group of modules (blue, red, black and yellow) took place at a mid-term between 30<sup>th</sup> to 90<sup>th</sup> minutes. The last group of modules (pink, purple, magenta, green and turquoise) has a common significant expression change between 90<sup>th</sup> and 120<sup>th</sup> minutes which implies the long-term response of yeast cells to the doxorubicin impulse (Figure 3.23).

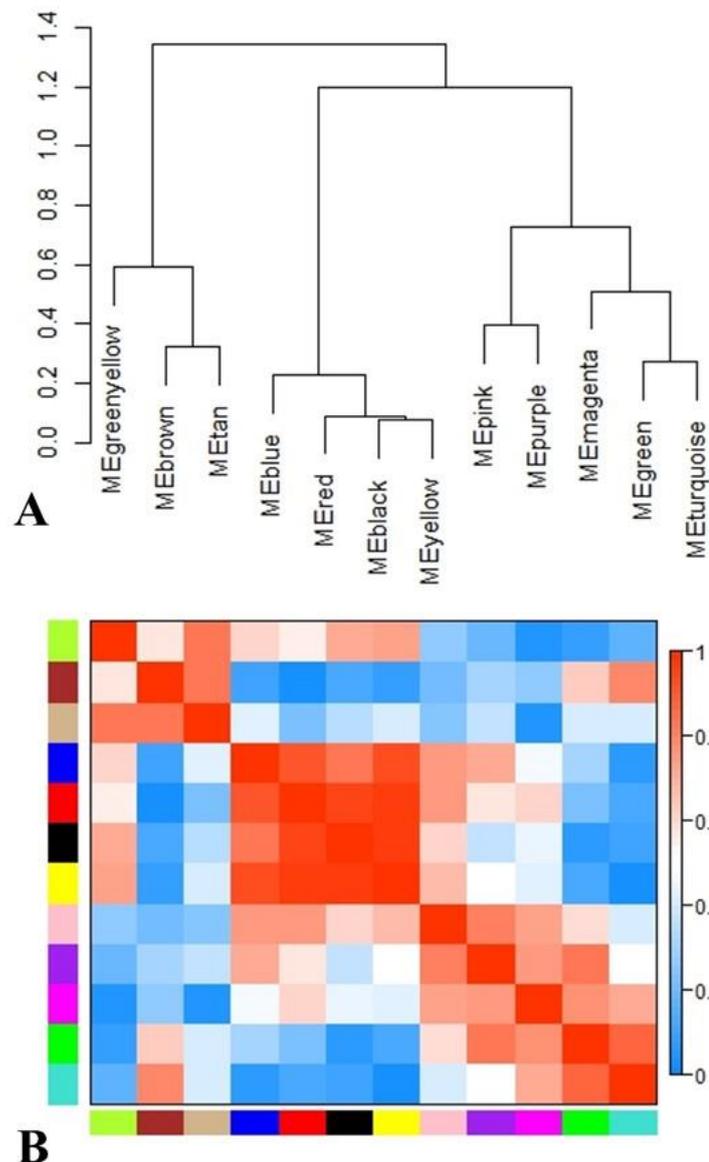


Figure 3.22. Meta-analysis of the modules identified by WGCNA for E-EDGE genes. Hierarchical organization of the modules (A) and heat-map of the modules (B).

All modules were found to be significantly enriched in more than one GO biological process term or pathway except modules ME9-magenta, ME11-greenyellow and ME12-Tan consisting of less than 20 genes.

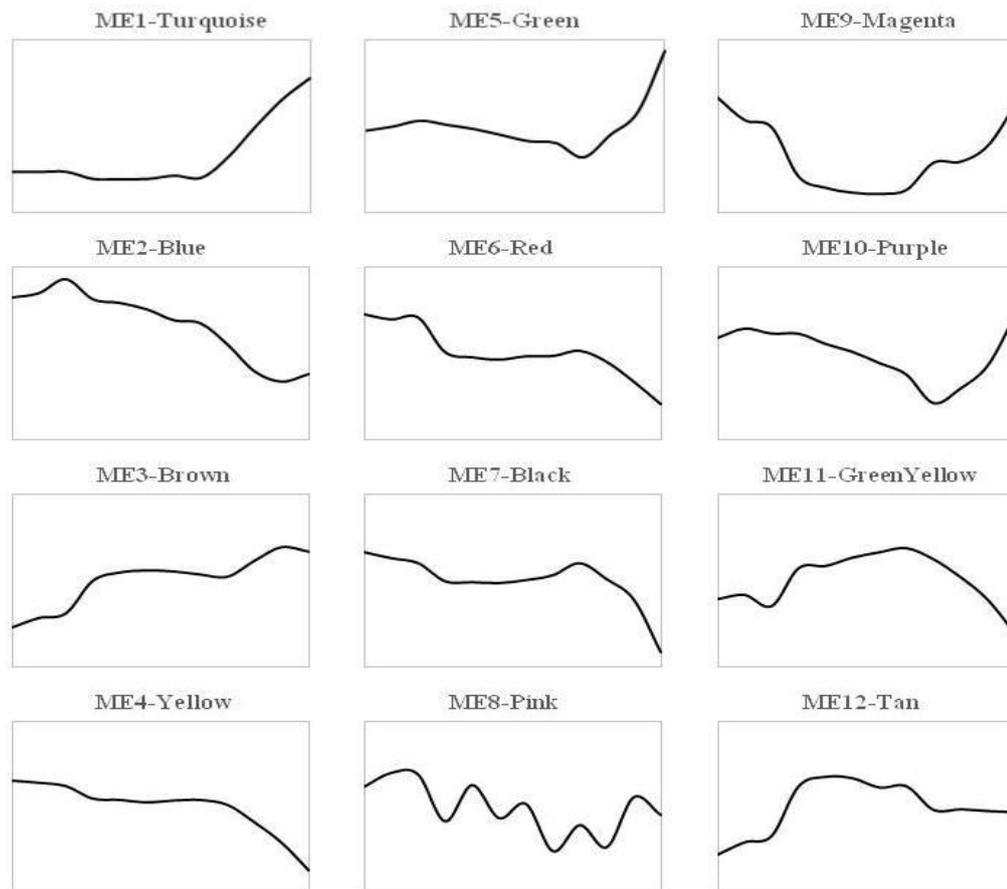


Figure 3.23. Eigengene profiles of the modules identified by WGCNA of genes identified by E-EDGE (x-axis is sampling points and y-axis is fold change).

The modules which are far from each other in the hierarchical eigengene tree (Figure 3.22) may include closely correlated genes. In addition to this, biological processes and pathways which are enriched in a module, identified by WGCNA, can not be interpreted as down-regulated or up-regulated without knowing whether they are positively or negatively correlated with the corresponding eigengene profile. In order to overcome these obstacles and to improve the content of the modules, genes having positive correlations with eigengenes were relocated using *signedKME* function of the WGCNA R package. A pipeline to reallocate genes was developed as described in Figure 3.24. For each gene PCCs with all eigengenes were ranked. If the highest PCC is over 0.75 then the gene was assigned to the

corresponding module. If the highest PCC is less than 0.75 then the lowest PCC was investigated. Provided that the lowest PCC is lower than -0.75 a new module was built as the counterpart of the module and the gene was assigned to the new module. If it is failed to satisfy both of the criteria then the gene was marked as outlier (Figure 3.24).

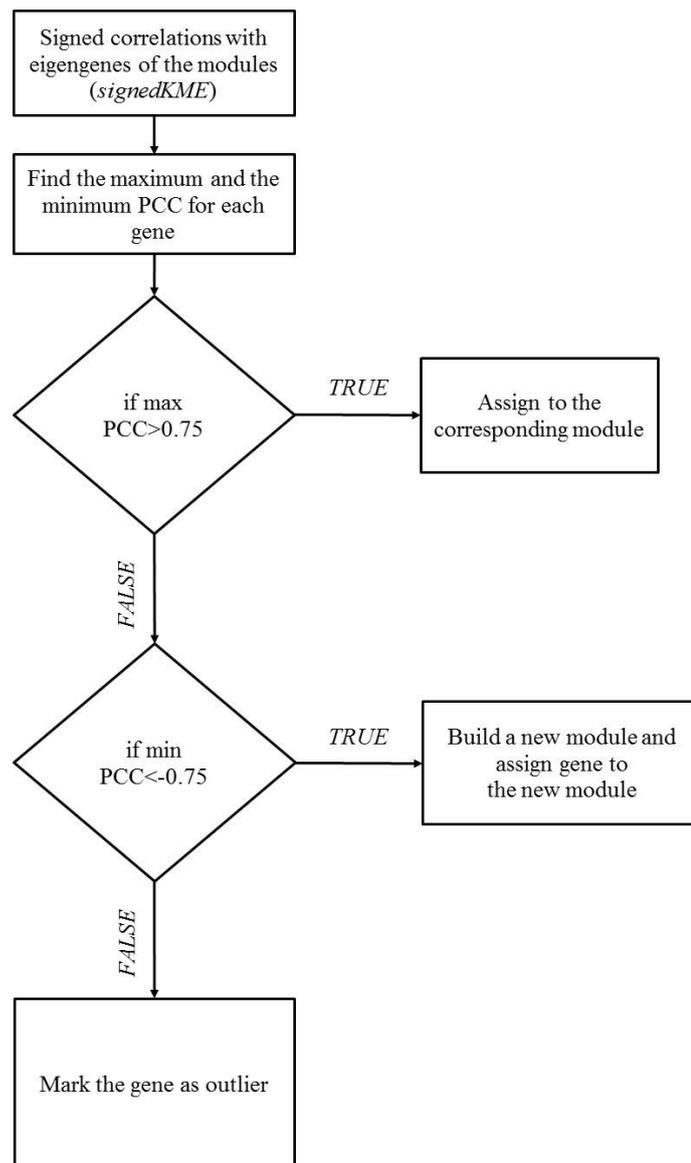


Figure 3.24. The pipeline of reallocation of the genes in the modules identified by WGCNA for E-EDGE genes.

Out of 2630 genes identified by E-EDGE 2464 genes were redistributed to the 12 modules with a positive correlation ( $PCC > 0.75$ ). New module IDs were indicated by an asterisk. Only one module which is negatively correlated with M2 has 10 or more genes and

identified as M2\*N. 98 nodes were found to be neither positively nor negatively correlated with any eigengene and identified as outliers (Table 3.10).

Table 3.10. Module ID's and sizes of E-EDGE genes determined by WGCNA after *signedKME* pipeline.

Module ID	Module Size	Module ID	Module Size	Module ID	Module Size
M1*	651	M6*	148	M11*	43
M2*	404	M7*	81	M12*	36
M3*	376	M8*	41	M2*N	28
M4*	530	M9*	27		
M5*	96	M10*	31		

Furthermore the average expression profiles of these modules may be used to identify the time dependent up- or down regulated biological processes or pathways in response to doxorubicin pulse (Figure 3.25).

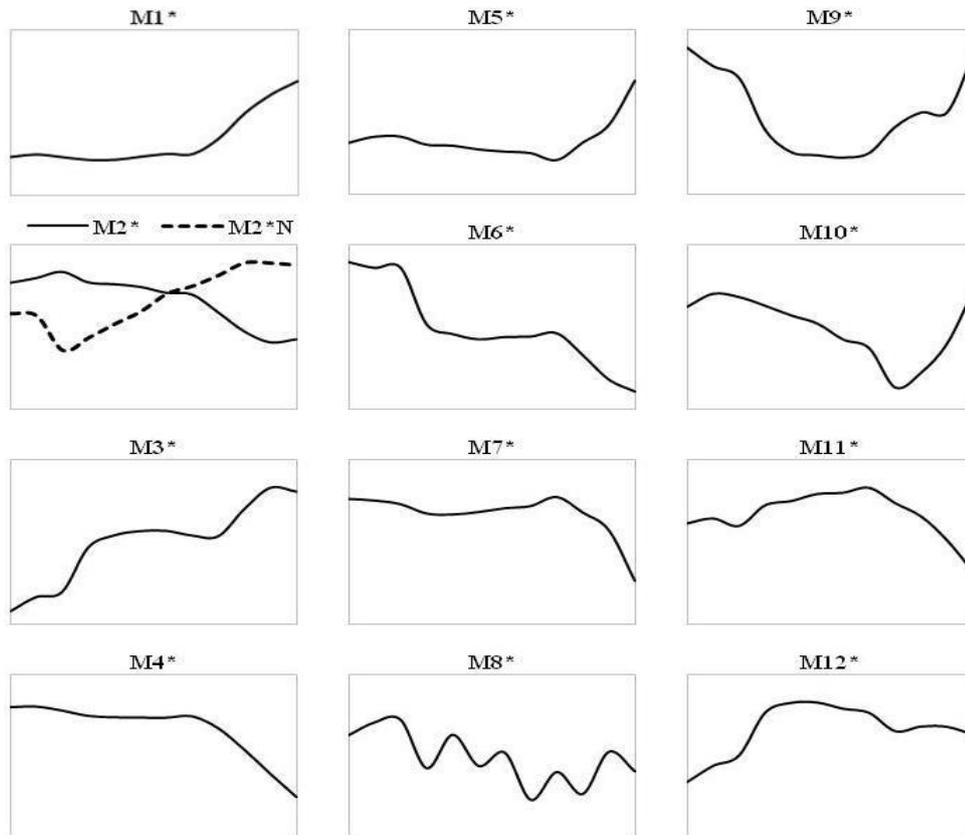


Figure 3.25. Average expression profiles of the positively correlated modules identified by WGCNA of the genes identified by E-EDGE (x-axis is sampling points and y-axis is fold change).

Genes which compose clusters M1\* and M5\* were shown to be induced after 30<sup>th</sup> and 60<sup>th</sup> minutes of the pulse respectively. M10\* displays a similar trend with M5\* in long term but it experiences a downregulation until the 60<sup>th</sup> minute (Figure 3.25). The biological process GO terms and pathways associated with these two cluster was given in Table 3.11.

Table 3.11. Selected GO biological process terms and pathways significantly associated with clusters M1\*, M5\* and M10\*.

<b>M1*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
protein localization in organelle	34	4.1E-05
siderophore transport	7	9.0E-05
ascospore wall assembly	13	1.6E-03
iron ion transport	11	1.7E-03
protein targeting to membrane	15	2.1E-03
mitochondrial translation	20	3.2E-03
protein import	25	4.7E-03
mitochondrial membrane organization	10	7.6E-03
mRNA 3'-end processing	11	7.7E-03
transcription initiation	15	8.4E-03
<b>M5*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ubiquitin-dependent protein catabolic process	19	1.9E-09
DNA-dependent DNA replication	11	2.5E-06
DNA repair	17	4.7E-06
protein refolding	6	8.0E-06
meiosis I	9	9.1E-05
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	14	1.4E-13
DNA replication	6	6.7E-04
Mismatch repair	5	9.8E-04
Base excision repair	4	7.7E-03
<b>M10*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
cell cycle phase	16	6.0E-10
mitotic sister chromatid segregation	8	1.3E-07
meiosis	6	3.0E-03
sexual reproduction	7	3.4E-03
regulation of DNA metabolic process	4	4.4E-03
cell cycle checkpoint	4	4.8E-03
regulation of cyclin-dependent protein kinase activity	3	5.3E-03
DNA replication	5	8.1E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Cell cycle	11	9.9E-10
Meiosis	7	2.3E-04

Major fold change in the expression levels of the genes identified in M2\* takes place between 30 to 120 minutes after the doxorubicin stimulus. Upregulation of the genes in M2\*N starts with a 5 minute delay after the pulse (Figure 3.25). The biological process GO terms and pathways associated with these two cluster was given in Table 3.12.

Table 3.12. Selected GO biological process terms and pathways significantly associated with clusters M2\* and its negatively correlated counterpart M2\*N.

<b>M2*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
response to osmotic stress	24	1.2E-06
positive regulation of transcription from RNA polymerase II promoter	20	6.3E-05
cell morphogenesis	26	9.9E-05
cell wall organization	36	1.4E-04
regulation of cell size	20	2.5E-04
nucleosome organization	14	3.3E-04
negative regulation of transcription from RNA polymerase II promoter	16	3.4E-04
endocytosis	18	4.8E-04
actin cytoskeleton organization	19	5.6E-04
response to organic substance	25	7.5E-04
intracellular signaling cascade	25	1.6E-03
cell surface receptor linked signal transduction	12	1.9E-03
chromatin assembly	9	2.2E-03
regulation of mitotic cell cycle	15	3.0E-03
regulation of meiosis	8	3.4E-03
establishment or maintenance of cell polarity	19	3.6E-03
chromatin modification	25	4.9E-03
chronological cell aging	7	6.6E-03
GMP biosynthetic process	4	6.7E-03
DNA repair	30	7.6E-03
mitotic sister chromatid segregation	13	9.9E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
MAPK signaling pathway	15	1.6E-05
Meiosis	18	6.1E-03
<b>M2*N</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
methionine biosynthetic process	4	4.2E-04
sulfate assimilation	3	9.4E-04
amine biosynthetic process	5	4.0E-03
serine family amino acid biosynthetic process	3	5.5E-03

Upregulation of M3\* cluster and downregulation of M6\* cluster occur in opposite two steps. Main changes in the expression levels of these genes take place between 5 to 10 minutes and 60 to 120 minutes after the pulse in the opposite directions (Figure 3.25). The

biological process GO terms and pathways associated with these two cluster was given in Table 3.13.

Table 3.13. Selected GO biological process terms and pathways significantly associated with clusters M3\* and M6\* which have opposite expression profiles.

<b>M3*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ribosome biogenesis	164	8.5E-109
ncRNA metabolic process	169	2.6E-105
RNA modification	56	1.2E-33
ribosomal large subunit assembly	27	3.1E-22
maturation of LSU-rRNA	13	1.4E-11
tRNA methylation	11	5.7E-08
transcription from RNA polymerase III promoter	13	9.1E-07
one-carbon metabolic process	16	3.4E-04
ribosomal large subunit export from nucleus	7	4.6E-04
regulation of translational initiation	6	1.2E-03
ribosomal small subunit assembly	6	3.5E-03
positive regulation of transcription from RNA polymerase I promoter	5	7.1E-03
RNA catabolic process	15	8.1E-03
RNA localization	18	8.1E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Ribosome	29	6.1E-11
RNA polymerase	15	1.6E-10
Pyrimidine metabolism	20	6.3E-09
Purine metabolism	19	2.8E-06
<b>M6*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
cellular response to heat	32	8.0E-20
vacuolar protein catabolic process	25	2.8E-17
cellular carbohydrate catabolic process	12	8.7E-07
glycoside biosynthetic process	5	8.5E-06
trehalose biosynthetic process	5	8.5E-06
alcohol catabolic process	9	9.7E-05
secondary metabolic process	7	2.3E-03
glycogen metabolic process	5	4.4E-03
autophagy	10	5.0E-03
glucose catabolic process	6	5.3E-03
nicotinamide metabolic process	6	5.3E-03
pyridine nucleotide metabolic process	6	6.2E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Starch and sucrose metabolism	11	9.8E-09

Genes which compose clusters M4\* and M7\* were shown to be downregulated after 30<sup>th</sup> and 60<sup>th</sup> minutes of the pulse, respectively. M11\* displays a similar trend with M4\* in long term but it experiences an induction until the 30<sup>th</sup> minute (Figure 3.25). The biological process GO terms and pathways associated with these two cluster was given in Table 3.14.

Table 3.14. Selected GO biological process terms and pathways significantly associated with clusters M4\*, M7\* and M11\*.

<b>M4*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
amine biosynthetic process	28	1.3E-04
cellular amino acid biosynthetic process	26	3.2E-04
translational elongation	50	5.9E-04
organic acid transport	17	2.6E-03
regulation of transcription, DNA-dependent	60	2.8E-03
lysine biosynthetic process via aminoadipic acid	5	3.4E-03
glycine metabolic process	5	5.7E-03
amine catabolic process	11	5.8E-03
glycerolipid metabolic process	16	8.3E-03
arginine biosynthetic process	5	8.9E-03
<b>M7*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
amine biosynthetic process	12	2.4E-06
arginine biosynthetic process	4	2.7E-04
histidine biosynthetic process	4	3.6E-04
ornithine metabolic process	3	3.6E-03
NAD metabolic process	4	4.4E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Histidine metabolism	5	4.1E-05
<b>M11*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
nitrogen compound biosynthetic process	8	7.5E-03

The genes involved in DNA repair, histone acetylation and meiosis associated were clustered in M8\* which displays the most noisy expression profile compared to the rest of the modules with the highest standard deviation over time. Genes in M9\* cluster are downregulated within the first 15 minutes and recovered in long-term, The genes involved

in growth within M12\* cluster were induced within the first 15 minutes (Figure 3.25). The biological process GO terms and pathways associated with these two cluster was given in Table 3.15. These clusters were not found to be associated with any pathway.

Table 3.15. Selected GO biological process terms significantly associated with clusters M8\*, M9\* and M12\*.

<b>M8*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
DNA repair	8	1.6E-03
histone acetylation	4	3.2E-03
meiosis	6	9.4E-03
M phase of meiotic cell cycle	6	9.4E-03
<b>M9*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
organic anion transport	3	6.3E-04
glycogen biosynthetic process	3	1.1E-03
vacuolar protein catabolic process	5	1.4E-03
response to temperature stimulus	6	1.9E-03
glucose metabolic process	4	8.5E-03
acetyl-CoA metabolic process	3	9.1E-03
<b>M12*</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
rRNA processing	15	4.1E-11
ribosome biogenesis	16	5.7E-10
tRNA modification	4	8.9E-03

All modules (M1\*- M12\* and M2\*N) were found to be significantly associated with at least one biological process GO term.

3.3.6.3. Regulation of the response to doxorubicin impulse. Expression profiles of 2580 genes out of 2630 genes identified by E-EDGE with known TF interactions were analyzed by DREM. Analysis of this subset by DREM revealed that doxorubicin pulse triggers complex regulatory events where 17 transcription factors (TFs) drives transcriptomic response into 10 branches associated with 80 to 463 genes at the end of time course (Figure 3.26).

Transcriptional profiles of the 2580 genes immediately splits into two branches A and B after doxorubicin pulse and A and B branches both underwent bifurcations after 1<sup>st</sup> minute yielding branches A1, A2, B1 and B2 respectively. After the second sampling point only B2 branch splits into B21 and B22. Until the 30<sup>th</sup> minute there were no splitting in the branches after that A2, B1 and B21 bifurcates again. The last bifurcations took place at 60<sup>th</sup> minute on A22 and B212 branches. Whole branches were coded in order to fulfill the same number of characters. The branch codes and number of genes they cover were tabulated with driving TFs (Figure 3.26) (Table 3.16).

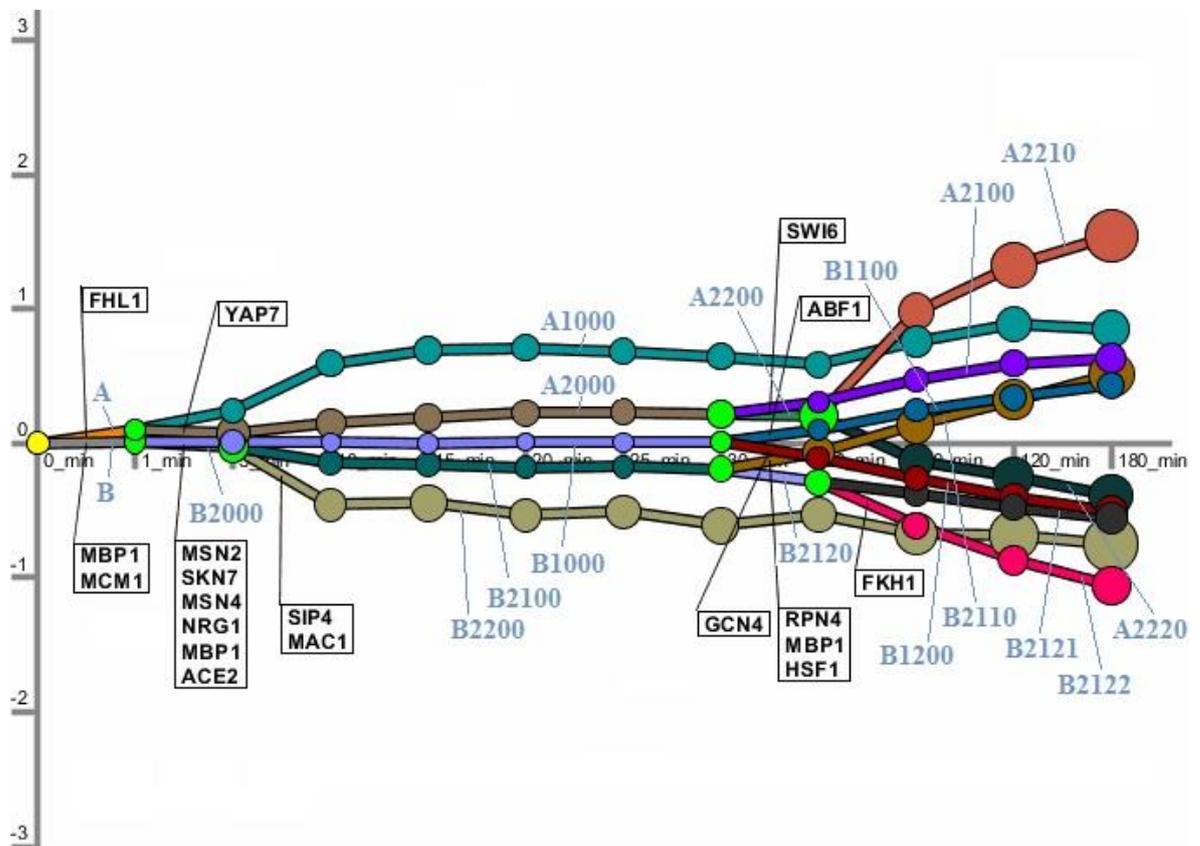


Figure 3.26. Dynamic regulatory map of differentially expressed yeast genes identified by E-EDGE in response to doxorubicin pulse. X-axis was arranged as uniform sample points.

Light-green dots indicate the location of bifurcations. Blue labels are the branch codes.

Immediate split in the transcriptional response was found to be mediated significantly by three regulators. A regulator of ribosomal protein (RP) transcription Fhl1p manages the

upregulated branch A associated with 769 genes. The branch B of 1811 genes which were downregulated immediately after doxorubicin addition were found to be significantly regulated by two transcriptional factors Mbp1p and Mcm1p which are involved in the regulation of cell cycle progression from G1 to S phase and cell-type-specific transcription and pheromone response, respectively (Figure 3.26).

Transcriptional levels of 548 genes regulated by a putative basic leucine zipper (bZIP) transcription factor; Yap7p remain steady for the first half hour after doxorubicin stress. 221 genes in the A2200 branch which displays relatively unchanged transcription under the control of Swi6p which regulates transcription at the G1/S transition, splits into two branches implicating a more drastic expression change at 60<sup>th</sup> minute (Figure 3.27).

Table 3.16. Number of genes in each branch and transcriptional regulators of the branches identified by DREM.

Minutes	2580 genes								
0	B (MBP1, MCM1) 1811 genes				A (FHL1) 769 genes				
1	B2000 963			B1000 848 genes (MSN2, SKN7, MSN4, NRG1, MBP1, ACE2)		A2000 (YAP7) 548 genes		A1000 221 genes	
5	B2100 756 genes								
10									
15									
20									
25	B2200 207 genes (SIP4, MAC1)								
30		B2120 529 genes		B2110 227 genes	B1200 (GCN4) 463 genes	B1100 (ABF1) 385 genes	A2200 (SWI6) 221 genes		
60		B2122 207 genes (FKH1)	B2121 322 genes	genes (RPN4, MBP1, HSF1)				A2100 327 genes	
90						A2220 141 genes	A2210 80 genes		
120									
180									

Six TFs (Msn2p, Skn7p, Msn4p, Nrg1p, Mbp1p and Ace2p) which have various regulatory roles in cell cycle, stress response and glucose repression, keep related expression profiles unchanged until 30<sup>th</sup> minute after the injection of the impulse. Up-regulation of 385

genes under the control of Abf1p which is involved in DNA replication and DNA repair and down-regulation of 463 genes under the control of Gcn4p which is an activator of amino acid biosynthesis occur 30 minutes after doxorubicin induction.

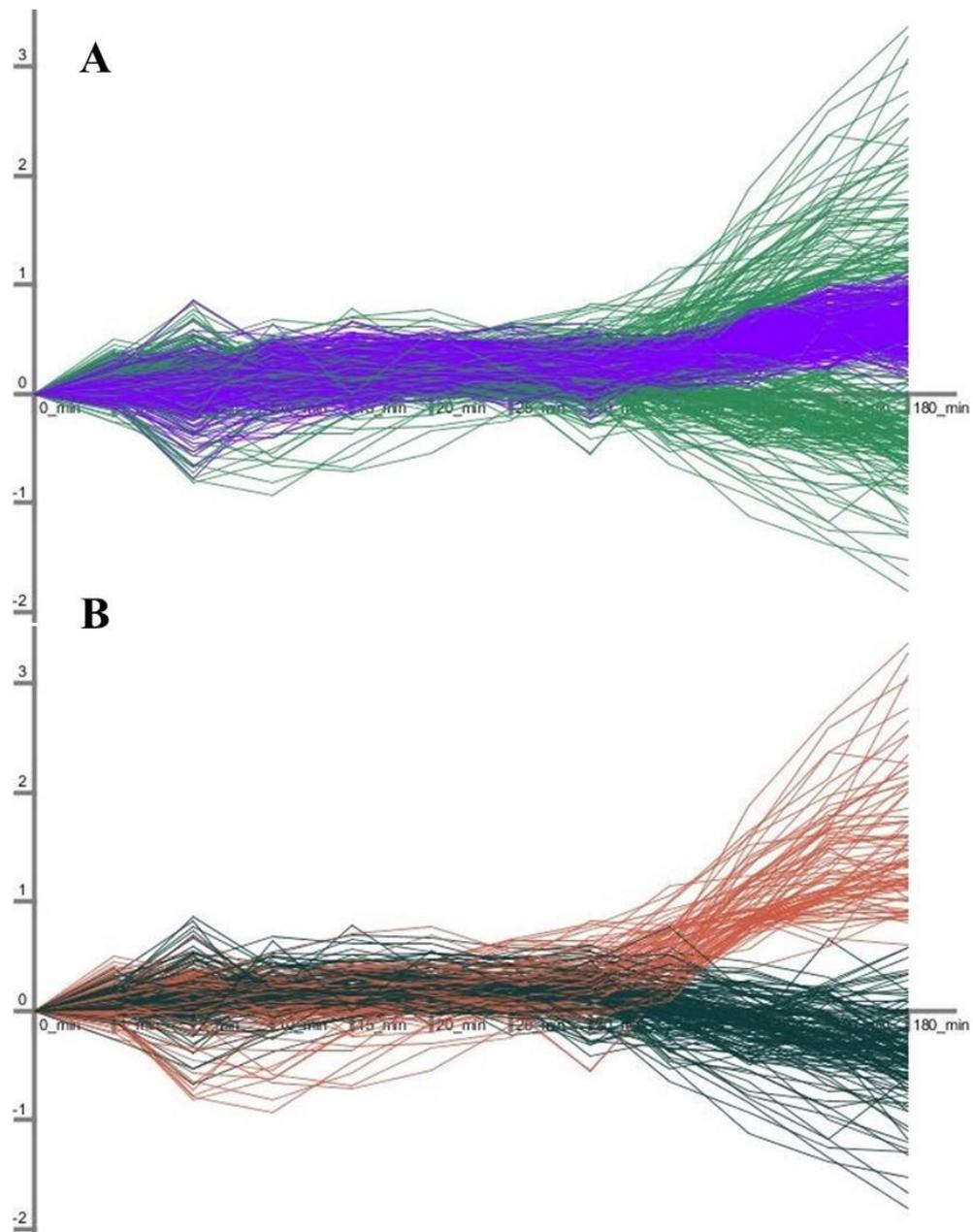


Figure 3.27. (A) Transcriptional profiles of A2100 (purple) and A2200 (green) branches and (B) two child branches of A2200; A2211 (brown) and A2212 (dark green).

B2100 splits into two branches. B2110 consists of 227 genes which were found to be induced at 30<sup>th</sup> minute under the control of heat shock transcription factor; Hsf1p, cell cycle regulator; Mbp1p and activator of proteasome genes; Rpn4p. 207 genes compose B2122

branch and were observed to be sharply down-regulated at 60<sup>th</sup> minute under the control of Fkh1p which is regulator of cell cycle. B2122 branch includes 207 genes which were regulated by copper-sensing Mac1p and Sip4p which is involved in positive regulation of gluconeogenesis. Mac1p is reported to be involved in regulation of genes required for high affinity copper transport and required for the regulation of yeast copper genes in response to DNA-damaging agents (Jungmann *et al.*, 1993; Dong *et al.*, 2013).

3.3.6.4. NP analysis of the genes identified by E-EDGE. Protein-protein interactions between proteins encoded by the genes identified by E-EDGE was extracted from the network which was used in section 3.3.3.1. The constructed network consists of 2630 proteins and 19,859 directionless edges without loops.

Pearson correlation coefficients (PCC) between expression profiles of the nodes connected by an edge was used to label the edge as positively correlated, negatively correlated or not correlated. The effect of the selected values of PCC on the frequency of possible gene pairs in the transcriptome of 2630 E-EDGE genes was analyzed in order to identify an acceptable threshold which can reduce the size of network to an amenable size (Figure 3.28).

PCC cut-off was selected to be more stringent as 0.85 and -0.85 for correlation and anti-correlation in NP analysis which correspond to the peaks in the Figure 3.28, respectively.

Resulting NP network for E-EDGE dataset consists of 1765 nodes and 5736 edges of which 4160 edges were found to connect positively correlated and 1576 edges to connect negatively correlated node couples. The previous NP network identified in the section 3.3.3.1 was about twice of the size of the new NP network considering the number of nodes and edges.

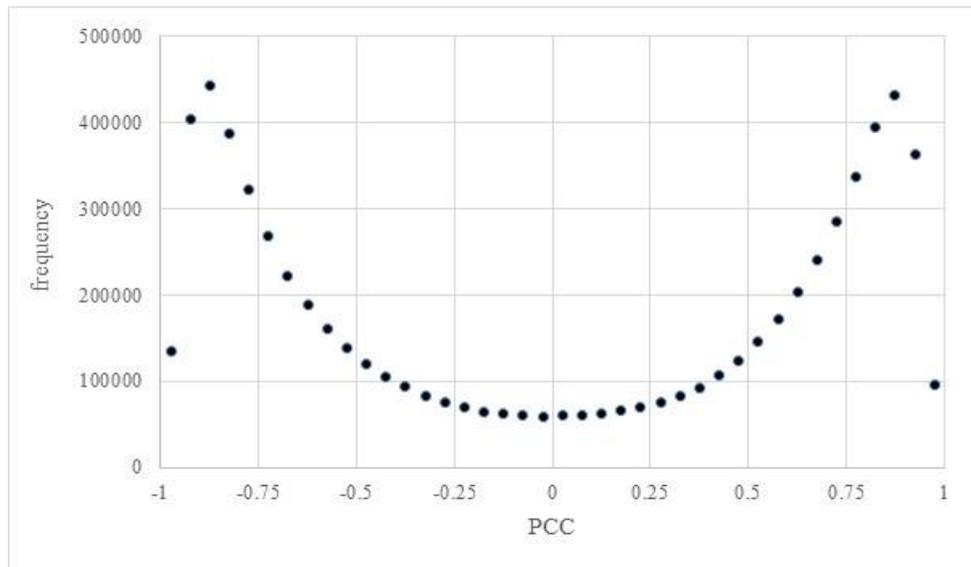


Figure 3.28. Frequency distribution of PCCs and network properties of NP network with respect to PCC threshold.

After hierarchical clustering of the expression profiles of the nodes present in NP network by HCE 3.0, modules with less than 1% intra-cluster anti-correlation edges were manually dissected. The network smoothly split into two modules without missing out any nodes and without violating the restriction on intra-cluster anti-correlation edges. Slightly larger one was named as K module with 890 nodes and the other was named as T module (Table 3.17).

Table 3.17. Distribution of the nodes and anti-correlated edges in NP network.

	K module	T module
Number of nodes	890	875
Percent of anti-correlated interactions within cluster	0%	0.07%
Percent of correlated interactions between clusters	0.13%	

These modules were further split into two sub-parts; Interface modules (KI and TI) consisting of nodes which interact with the opposite module and core modules (KC and TC) consisting of nodes which has only intra-modular interactions. 47 nodes in NP network is disconnected from bulk of the network.

Average expression profiles of the four modules identified by NP analysis were also presented in Figure 3.29. Average expression profiles for the core and the interface of each module are almost identical. KC and KI modules display a downregulated trend similar to the cluster M6\* which was identified by WGCNA. Genes in TC and TI modules which were oppositely expressed with genes in KC and KI, have an expression profile correlated with cluster M3\* which was also identified by WGCNA.

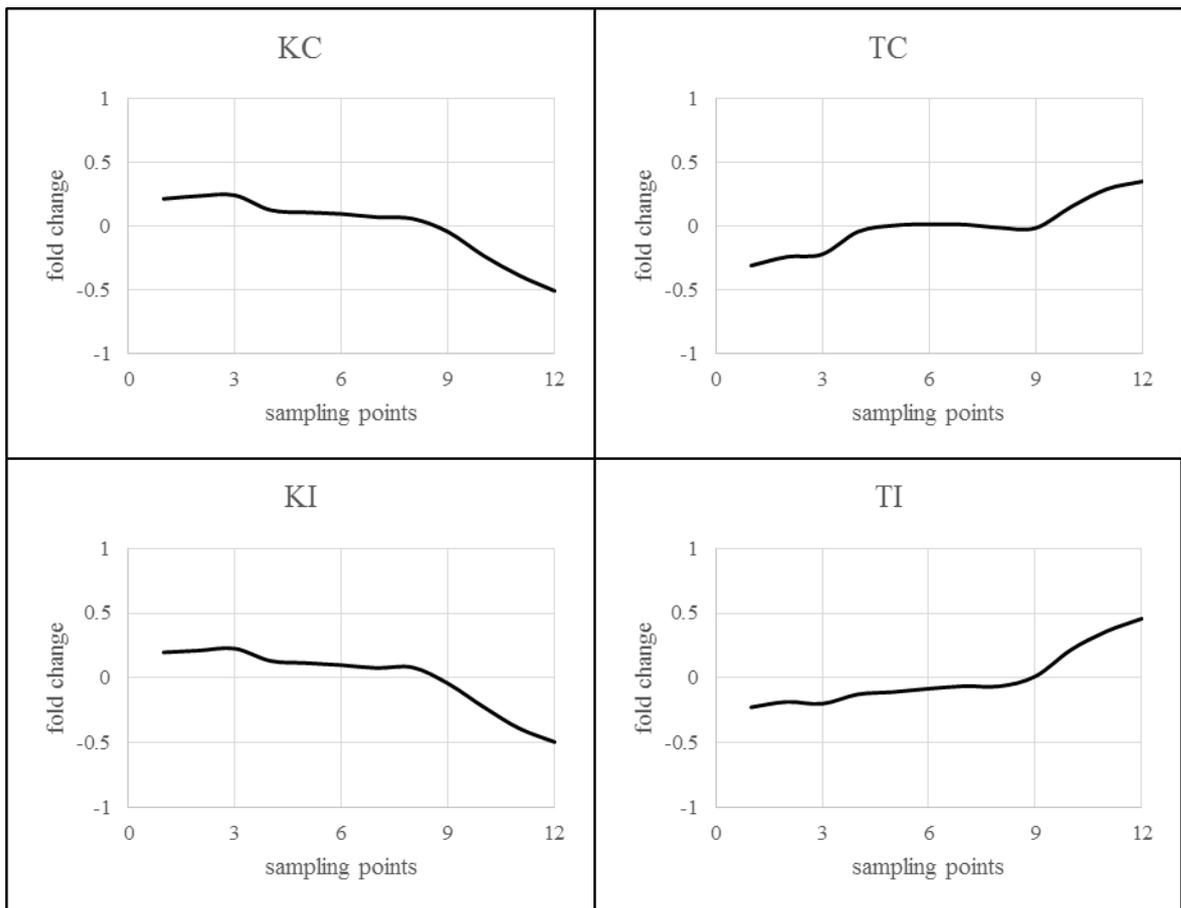


Figure 3.29. Average expression profiles of the modules identified by NP analysis of E-EDGE genes

Functional enrichments of the TC, TI, KI and KC modules were also investigated in terms of GO biological process term and pathway enrichments.

297 nodes which constitute KC module were found to be enriched in proteasomal ubiquitin-dependent protein catabolic process, DNA-dependent regulation of transcription, cytokinetic process and negative regulation of gluconeogenesis. Larger sub-module of

downregulated K module is the interface module KI which includes 593 nodes. GO biological processes which were found to be significantly associated with the nodes of interface module KI are much more diverse; from stress responsive terms to glucose metabolism, amino acid metabolism and biological processes related to regulation and transport (Table 3.18.).

Table 3.18. Selected GO biological process terms significantly associated with KC and KI modules.

<b>KC</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
proteasomal ubiquitin-dependent protein catabolic process	13	5.3E-04
regulation of transcription, DNA-dependent	40	5.8E-04
cytokinetic process	12	7.8E-03
negative regulation of gluconeogenesis	4	8.2E-03
<b>KI</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
response to osmotic stress	27	4.7E-05
vacuolar protein catabolic process	27	3.7E-04
protein amino acid phosphorylation	30	4.8E-04
cellular response to stress	86	1.3E-03
small GTPase mediated signal transduction	20	1.3E-03
phosphatidylcholine biosynthetic process	6	1.5E-03
alcohol catabolic process	17	1.6E-03
cell division	57	2.2E-03
glucose catabolic process	14	2.6E-03
lysine biosynthetic process	6	2.6E-03
negative regulation of transcription, DNA-dependent	33	2.7E-03
regulation of cell cycle process	22	2.8E-03
purine nucleotide biosynthetic process	18	3.0E-03
cellular carbohydrate catabolic process	18	3.5E-03
glycoside biosynthetic process	5	4.1E-03
trehalose biosynthetic process	5	4.1E-03
intracellular signaling cascade	32	4.2E-03
drug transport	8	4.4E-03
nucleosome organization	15	5.2E-03
aging	17	6.4E-03
amine biosynthetic process	27	6.8E-03
glucose transport	5	7.4E-03
chromatin assembly	10	8.2E-03
positive regulation of transcription from RNA polymerase II promoter	20	9.6E-03

Core sub-module of upregulated T module, TC, which consists of 346 nodes were found to be linked to growth related biological processes like ribosome biogenesis. Wider range of pathway enrichments were identified for the TC module compared to its biological process term enrichments. Ribosome pathway, proteasome, mismatch repair and DNA replication pathways were also found to be related to TC module. Similar to the case in K module, interface of the T module has more diverse biological process attachments. 529 nodes which reside in the interface module TI were found to be significantly enriched in biological processes like ribosome biogenesis, protein folding, transcription initiation, siderophore transport, regulation of translational fidelity, RNA localization, protein targeting to membrane, vacuolar acidification, ascospore wall assembly and proteasome pathway (Table 3.19).

Table 3.19. Selected GO biological process terms and pathways significantly associated with TC and TI modules.

<b>TC</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ribosome biogenesis	117	2.1E-57
ncRNA metabolic process	116	1.8E-50
ribosomal large subunit export from nucleus	7	3.3E-04
rRNA modification	7	3.3E-04
rRNA catabolic process	8	6.1E-04
tRNA methylation	6	4.8E-03
positive regulation of transcription from RNA polymerase I promoter	5	5.7E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Pyrimidine metabolism	18	4.6E-07
RNA polymerase	11	5.1E-06
Purine metabolism	17	8.1E-05
Ribosome	17	3.0E-03
Proteasome	8	5.3E-03
Mismatch repair	6	5.4E-03
DNA replication	7	9.9E-03
<b>TI</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ribosome biogenesis	71	8.4E-10
protein folding	29	5.3E-07
transcription initiation	17	3.9E-04
snoRNA 3'-end processing	9	6.7E-04
protein complex assembly	40	7.2E-04
siderophore transport	6	7.4E-04

Table 3.19. Selected GO biological process terms and pathways significantly associated with TC and TI modules. Cont.

RNA polyadenylation	9	1.7E-03
macromolecular complex disassembly	18	2.0E-03
mRNA 3'-end processing	11	3.8E-03
regulation of translational fidelity	6	5.5E-03
RNA localization	23	6.6E-03
protein import into nucleus	13	7.1E-03
protein targeting to membrane	13	7.1E-03
cellular macromolecular complex assembly	44	7.4E-03
vacuolar acidification	8	7.7E-03
ascospore wall assembly	11	7.8E-03
termination of RNA polymerase II transcription	7	9.4E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	11	1.6E-03

### 3.4. Discussion

#### 3.4.1. Dynamic response of Yeast cells to doxorubicin pulse

The clustering and analysis of the transcriptomic response of *S. cerevisiae* cells to doxorubicin indicated that the genes involved in DNA replication, mismatched repair, cell cycle and base excision repair pathways were found to remain without any change in their expression levels within the first five minutes after the induction of doxorubicin and they were down-regulated until the 60<sup>th</sup> minute where they start to be up-regulated.

These results are in correlation with the fact that doxorubicin interacts with DNA through intercalation like other anthracycline antibiotics leading to DNA damage (Fornari *et al.*, 1994; Anders *et al.*, 2013). It is also known that during the transcription doxorubicin inhibits topoisomerase II which has an important role in relieving torsional stress during replication and transcription by forming double stranded breaks (Pommier and Marchand 2012).

One set of genes involved in proteasome also behaved similarly. However the expression levels of another set remained unchanged for 60 minutes and then up-regulated. Selective inhibition of proteasomes in cancer cells is an anticancer treatment strategy whose

efficacy lies in blocking metabolic functions, inducing apoptosis, and sensitizing malignant cells and tumors to chemotherapeutic agents and radiation (Voorhees and Orlowski 2006; Orlowski and Kuhn 2008; Yang *et al.* 2009). Loss of 20S proteasome activator, *BLM10* was reported to be downregulating genes important for genomic integrity, increasing DNA damage and increasing sensitivity to doxorubicin (Doherty *et al.*, 2012).

Another set of genes whose expression levels were observed to be down regulated after 30 minutes consist of genes involved in meiosis, regulation of cell cycle, cell wall organization, response to osmotic stress and chromosome organization.

WGCNA analysis of differentially and significantly expressed genes has also shown the up-regulation of the genes involved in DNA repair, DNA replication base excision repair and proteasome after 60<sup>th</sup> minute and supported these observations. Cell cycle phase, meiotic cell division, cell cycle check-point biological processes were affected from the beginning and the genes involved in this processes were down regulated until 60<sup>th</sup> minute and then up-regulated. Proteasome related genes were verified to be remained constant in their expression levels within the first five minutes after the induction of doxorubicin and they were down-regulated until the 60<sup>th</sup> minute where they start to be up-regulated.

DREM analysis of differentially and significantly expressed genes also indicated that Mbp1p and Mcm1p which are involved in the regulation of cell cycle progression from G1 to S phase and cell type specific regulation and pheromone response were the main regulators of the down-regulated genes at 5<sup>th</sup> minute of the doxorubicin application. It is believed that doxorubicin prevents proliferation of cancer cells through inhibiting topoisomerase enzyme which blocks DNA replication and cause cell cycle arrest at S phase (Ichikawa *et al.*, 2014).

Rpn4p, activator of proteasome genes, was found to be upregulated one set of genes after 30<sup>th</sup> minute of the impulse. Relative distribution of Rpn4p to the nucleus is also known to be increased by DNA replication stress. DREM has also identified Abf1p which involves in DNA replication and repair, as a regulator of a group of upregulated genes after 30<sup>th</sup> minute of the pulse.

NP analysis has also indicated the verified upregulation of the genes involved in mismatch repair and DNA replication pathways after 60<sup>th</sup> minute. Genes which are significantly associated with proteasome pathway were also identified to be upregulated after 60<sup>th</sup> minute.

Induction of ribosome biogenesis and coherent down-regulation of vacuolar protein catabolic process were found to be overrepresented in upregulated and downregulated clusters, respectively. Up-regulation and down-regulation of these genes took place immediately after doxorubicin addition to the chemostat culture. All of the applied methods revealed similar results in correlation with these observations.

Interestingly drugs were reported to inhibit ribosomal RNA synthesis either at the level of (i) rRNA transcription (*e.g.* oxaliplatin, doxorubicin, mitoxantrone, methotrexate), (ii) early rRNA processing (*e.g.* camptothecin, flavopiridol, roscovitine), or (iii) late rRNA processing (*e.g.* 5-fluorouracil, MG-132, homoharringtonine) (Burger *et al.*, 2010). It has been also reported that genotoxic stresses can cause cell cycle arrest and then halt rRNA synthesis as a secondary effect in long term (Tsai and Pederson, 2014). Time interval of the chemostat experiment implemented in this study might not be enough to capture this expected inhibition of ribosome biogenesis.

Analysis of the expression profiles by SOMS indicated that the expression levels of the genes involved in siderophore transport remained unchanged until one hour before sharply up regulated and the genes involved in cellular iron ion homeostasis were found to be up-regulated after 5<sup>th</sup> minute.

WGCNA indicated that the genes involved in siderophore transport, iron ion homeostasis, mitochondrial translation, mitochondrial membrane organization and transcription initiation biological processes were also found to be up-regulated 30 minutes after the pulse.

NP analysis was also verified that siderophore transport related genes were up-regulated after doxorubicin impulse with a one hour delay.

Doxorubicin was reported to be a powerful iron chelator (Myers, 1998). It might be due to the iron arrest caused by doxorubin that siderophore transport linked genes were found to be induced in response to the doxorubicin impulse. In accordance with these observations it was also reported that exposure of bovine aortic endothelial cells to doxorubicin cause apoptosis accompanied by induction of iron uptake (Kotamraju *et al.*, 2002).

The cell cycle takes place normally if deoxyribonucleotides are available. The process is controlled by ribonucleotide reductase and thioredoxin (Hashemy *et al.*, 2006). In *S. cerevisiae* thioredoxin and the metabolic pathway it is involved are regulated with transcription factor Yap1p (Colleman *et al.*, 1999; Delaunay *et al.*, 2002; Moye-Rowley, 2003). Doxorubicin induces oxidative stress, which prolongs the G1 phase. However, for unknown reasons, doxorubicin increases GSH concentration in *S. cerevisiae* cells, probably due to the deregulation of deoxyribose synthesis, which, as we believe, is to block the cell cycle in the S phase. Another reason could be that oxidative stress was reported to enhance expression of sulfur assimilation genes (Riboldi *et al.*, 2014) and consequently GSH, cysteine and methionine biosynthesis.

SOMs and WGCNA have verified that methionine biosynthetic process and sulfate assimilation processes were upregulated between 5<sup>th</sup> and 90<sup>th</sup> minutes. The genes involved in sulfur amino acid biosynthetic process, biotin and pyridoxine biological processes were also found to be affected similarly. Interestingly GSH biosynthesis associated genes were found to be down-regulated after 60<sup>th</sup> minute. It could be a response to the accumulation of GSH which was caused by doxorubicin.

The expression levels of genes involved in arginine and histidine biosynthetic processes, ornithine and glycerophospholipid metabolic processes were observed slightly reduced during the first 5 minutes, then remain constant before to be repressed after 60 minutes sharply. WGCNA has also provided additional supportive information about the down-regulation of the genes involved in arginine and histidine biosynthesis after 60<sup>th</sup> minute of the impulse.

Individual analysis have revealed that drug transporters *PDR5*, *AQR1*, *QDR3*, *SNQ2* and *ATRI* were found to be downregulated after 30<sup>th</sup> to 60<sup>th</sup> minutes of the pulse but the process could not identified to be significantly associated with any identified cluster.

### **3.4.2. Improvements provided by E-EDGE**

An extended version of EDGE was proposed in order to identify differentially expressed genes out of a dynamic dataset. Clustering approaches were applied to this subset assuming that biologically irrelevant or inactive data was eliminated which can improve biological associations of the resulting modules or clusters.

Although size of the subset identified by E-EDGE is more than twice the size of the subset identified by variance filtering in SOMs analysis, biological significance of the resulting clusters were improved. Even after manual merging of the clusters which were identified in the previous SOMs analysis there were unannotated meta-clusters like MC6. SOMs of the genes identified by E-EDGE yielded nine clusters with distinct biological roles.

WGCNA of whole transcriptome resulted in three modules without any GO biological process term enrichments. These modules have sizes of 40 to 70 nodes. The subset identified by E-EDGE is comprised of less than half of the nodes which consists whole set but still the largest un-annotated module identified by subordinated WGCNA was the one with 20 genes indicating the success in extracting the significant subset.

PCC distribution of the genes identified by E-EDGE (Figure 3.28) scatters in a different way compared to that of whole dataset (Figure 3.10). The difference is raised from the low frequency of the not correlated gene pairs in the E-EDGE set which is an indicator of the integrity of the subset without noise.

A comparison of TFs identified by DREM analysis of whole transcriptome data and the genes identified by E-EDGE revealed 8 common TFs among the two datasets. Analysis of the genes identified by E-EDGE revealed some of the stress responsive TFs (*Msn2p*, *Msn4p* and *Mcm1p*) and regulators of the processes like proteasome (*Rpn4p*), glucose metabolism (*Sip4p* and *Nrg1p*), copper-sensing (*Mac1p*) and ribosome biogenesis (*Fhl1p*)

which could not be identified when the global transcriptome was used (Figure 3.12 and Figure 3.26).

All these findings suggest that identifying differentially expressed genes by E-EDGE and then implementing clustering analysis on the resulting significant subset of the data can strengthen the results and unveil additional biological data.

### **3.4.3. Improvements provided by the reallocation pipeline attached to WGCNA**

A reallocation pipeline was also developed and proposed to improve WGCNA. The reallocation yielded stronger biological attachments. 91.7% of the GO biological process terms out of 156 terms which have been identified before and after allocation has stronger attachments to the modules identified by proposed reallocation in terms of p-values.

## **3.5. Future prospects**

Up-regulation of ribosome biogenesis and repression of arginine biosynthesis are needed to be investigated in more detail. These unexpected outcomes of the experiments should be verified and underlying reasons should be clarified by using deletion strains.

The magnitude of the transcriptional response could be increased by repeating the chemostat experiments with higher doses of doxorubicin pulses.

The pipeline developed in this study can be generalized for dynamic data analysis by implementing for other dynamic datasets. The improvements provided by E-EDGE approach and re-allocation process in WGCNA can also be investigated independently.

Topological analysis of the weighted co-expression network which was constructed by WGCNA can be investigated further by using WSL-EC. Topological centrality in co-expression networks and its biological implications may unveil additional information.

## 4. INVESTIGATION OF DYNAMIC RESPONSE OF YEAST CELLS TO DNA DAMAGE IN TRANSCRIPTOMIC LANDSCAPE

### 4.1. Background

DNA damage can occur in many ways like damage on bases, mismatch of bases, single-strand or double-strand breaks upon environmental stresses. Distinct stress agents may cause different damage types and there are distinct DNA repair mechanisms in order to fix these damages (Friedberg *et al.*, 2005).

DNA damage caused by chemical and physical stresses may lead to cancer and degenerative diseases (Workman *et al.*, 2006). Cells which suffer DNA damage triggers mechanisms regulating transcriptional response of diverse biological processes other than DNA repair like cell cycle, apoptosis, catabolic processes and ribosome biosynthesis (Harper and Elledge, 2007; Shalem *et al.*, 2008; Ciccia and Elledge, 2010).

*Saccharomyces cerevisiae* is well defined and easy to manipulate model organism. As a matter of fact until the last two decades the most of the information on DNA damage response was coming from yeast (Harper and Elledge, 2007). Early knowledge was solely around DNA repair genes, and their interactors but high throughput transcriptional assays revealed many more genes with diverse functions response to DNA damage in transcriptional level (Jelinsky and Samson, 1999). For instance Rpn4p, which is a proteasome-associated protein, was reported to be linked to base excision mechanism as well (Jelinsky *et al.*, 2000).

Mec1p which is homolog of human ATR kinase plays a crucial role in sensing and signaling DNA damage in yeast. DNA damage signal triggers Chk kinases (Chk1p, Rad53 and Dun1) and Chk kinases, in turn, activates various mechanisms like DNA repair machinery and cell cycle arrest (Gasch *et al.*, 2001; Friedberg *et al.*, 2005).

Carcinogen induced genotoxicity analysis of haploid *S.cerevisiae* deletion strains revealed that additional proteins of non-nuclear protective pathways in vacuole or endosome modulate toxicity other than known DNA repair proteins (Begley *et al.*, 2004).

Static transcriptome data collected after exposure of yeast to methyl-methanesulfonate (MMS) unveiled 30 DNA-damage related transcriptional factors (TF) out of which *YAP1*, *ACE2*, *ASH1* and *SWI5* were identified to be differentially expressed while either deletion strains or targets of the rest of the TFs were identified to be sensitive to exposure to MMS. These DNA-damage related TFs were reported to be regulators of biological processes like cell cycle, pyrimidine metabolism, ribosome biogenesis, response to heat and redox homeostasis (Figure 4.1) (Workman *et al.*, 2006). Differentially expressed *YAP1* is required for oxidative stress tolerance and the other three TFs; *ASH1*, *ACE2* and its paralog *SWI5* involve in regulation of G1/S transition in the mitotic cell cycle.

In a wide study, 80,000 double mutant yeast strains were treated with MMS in order to construct static genetic maps of each condition; with or without MMS. More than half of the interactions within the genetic networks were reported to be condition specific. In the light of differential network analysis it was suggested that *SLT2*, *BCK1* and *CBF1* might have roles in DNA damage response (Bandyopadhyay *et al.*, 2010). It has been also reported that *TELL1*, *MEC1*, *LCD1*, *RAD9*, *DPB11*, *MRC1*, *CHK1* and *RAD53* play important roles in DNA damage signaling of yeast (Polo and Jackson, 2011).

Dynamic transcriptional response of yeast cells to MMS were also reported but the focus of the work was on mRNA degradation rather than DNA damage. It was reported that proteolysis related genes were induced and stabilized, growth related genes were repressed and destabilized and RNA processing related genes were repressed and stabilized (Shalem *et al.*, 2008).

Another dynamic transcriptome analysis regarding DNA damage was carried out in order to investigate effect of UV exposure on yeast cells. This study suggests that *SNF1* and

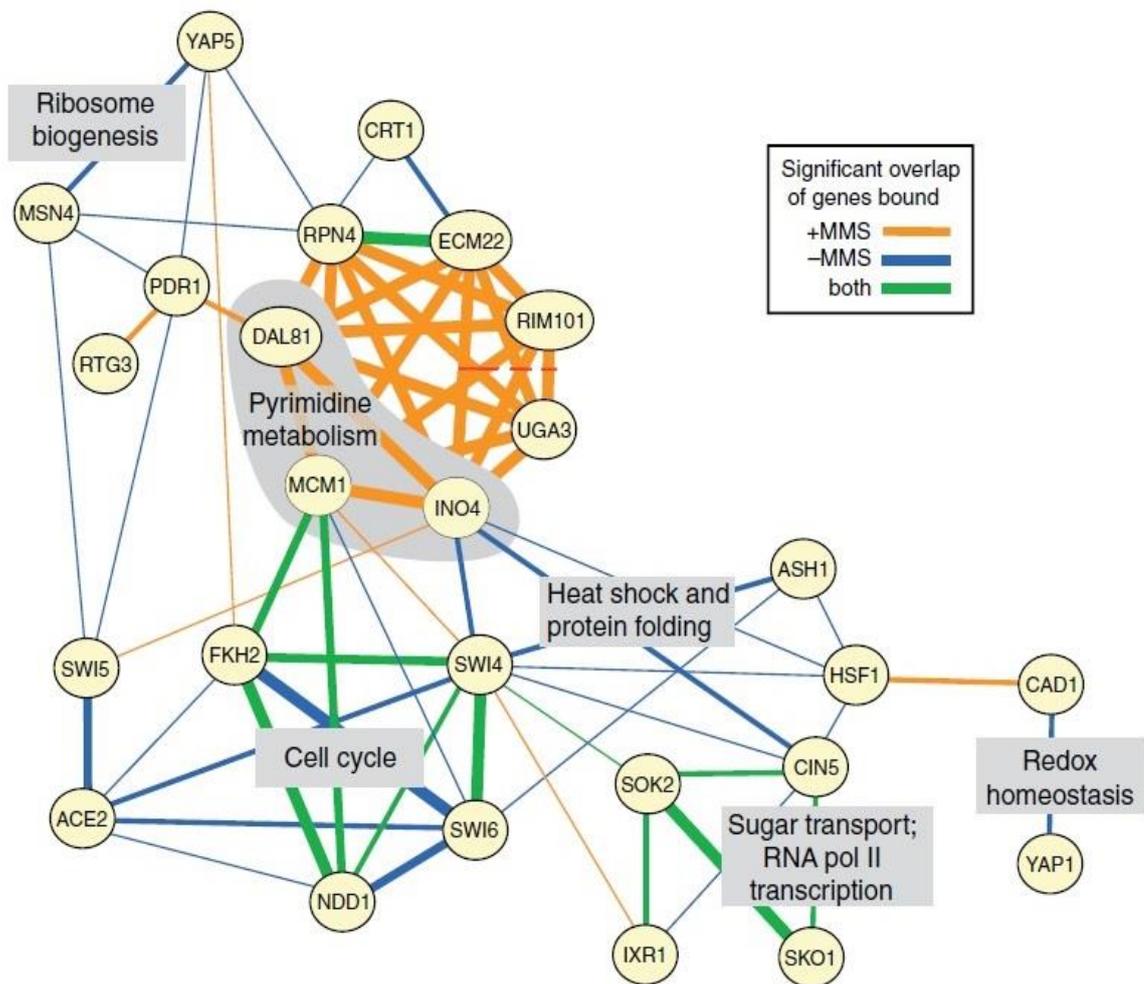


Figure 4.1. TF network of overlapping targets in response to MMS (Workman *et al.*, 2006).

*RAD23* which has no previously known regulatory control role regulate UV-responsive gene expression. It was identified that oxidative stress response and amino and carboxylic acid metabolism processes related genes were up-regulated upon UV-exposure while ribosome biogenesis and cell cycle regulation like processes were repressed (Wade *et al.*, 2009).

In this study time series microarray datasets collected after a genotoxic stress in *Saccharomyces cerevisiae* were selected from the literature. Selected datasets were analyzed by well-established clustering methods and integrative approaches. Differentially expressed genes were identified by E-EDGE approach which was defined in the previous section.

## 4.2. Methods

### 4.2.1. Transcriptome Data

Two data sets were selected from literature where the DNA damaging agent was methyl methane sulfonate (MMS) (Shalem *et al.*, 2008) (1-P) (GSE12222) and exposure to UV (Wade *et al.*, 2009) (2-WT) (GSE16799).

### 4.2.2. Transcriptome Data Analysis

The dataset 1-P consists of six time points while the dataset 2-WT has four time points. The two datasets were added to have a single transcriptome data of 10 time points. Each of these two datasets was normalized before integration in order to avoid possible bias which might rise from differences in experimental conditions. The resulting dynamic transcriptome data was analyzed by following the pipeline which formed in the previous chapter (Figure 4.2).

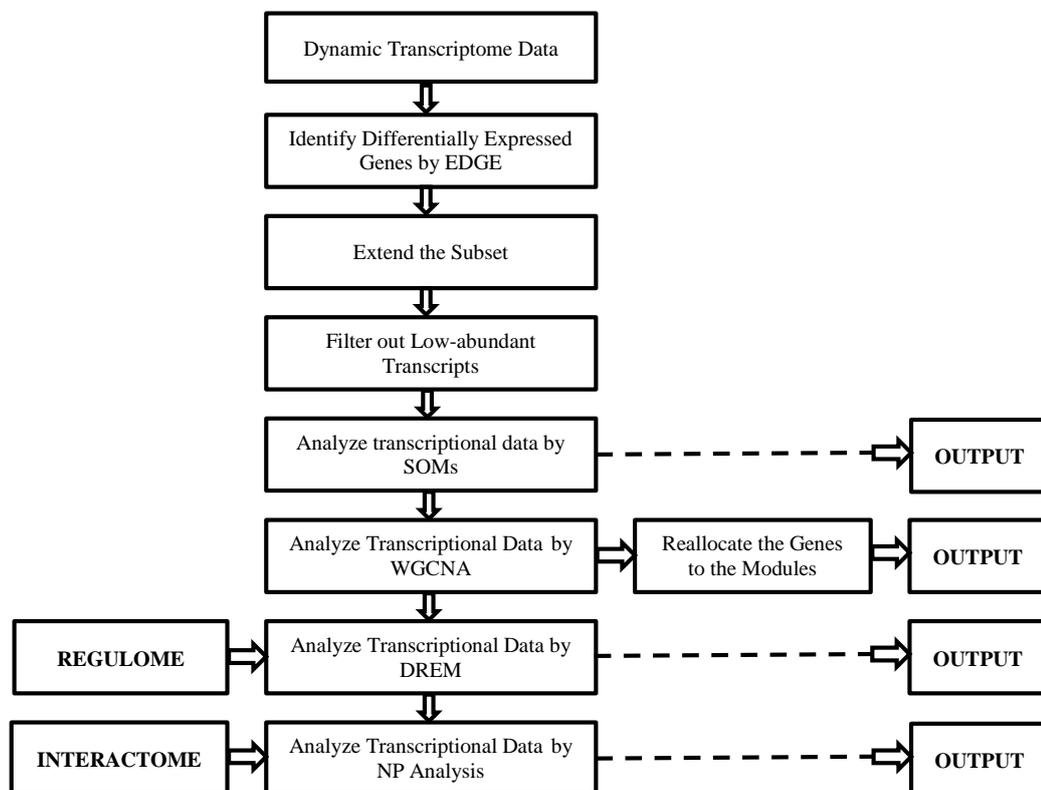


Figure 4.2. The procedure for dynamic transcriptional data analysis

Starting with identification of differentially and significantly expressed genes by E-EDGE and continuing with SOMs, WGCNA, reallocation in WGCNA, DREM and finally NP analysis. Only in NP analysis two dataset were analyzed separately (Figure 4.2).

SOMs, WGCNA, E-EDGE, DREM, NP analysis and functional annotations were carried out as described in the Section 3.2.3.

### 4.3. Results

Shalem and colleagues used *S. cerevisiae* strain Y262 and grew cells in YPD medium. After treating cells in batch culture by 0.1% MMS samples were collected at 0, 30, 60, 100, 140 and 180 minutes and mRNA abundance were measured by microarray technology. Fold-changes were used to determine the significantly and differentially expressed genes (Shalem *et al.*, 2008). The reported study focused on mRNA degradation and shortly analysed dynamics of the DNA damage response.

Wade and colleagues used *Saccharomyces cerevisiae* strains derived from YPH499 and grew cells in YPD medium. Cells were irradiated with 100 J/m<sup>2</sup> and samples were collected at 0, 15, 30 and 60 minutes for RNA isolation. Hierarchical clustering based on Spearman correlation was used to analyze the data. The focus of the study was on differences between transcriptomic responses of wild type and deletion strains (*snf1Δ* and *rad23Δ*) to UV irradiation (Wade *et al.*, 2009).

In the present study we merged these two datasets, identified differentially expressed genes, clustered the dataset and integrated with different omics in order to get a deeper understanding of the dynamics of transcriptional response to DNA damage induced by treatment with MMS and exposure to UV.

#### 4.3.1. Identification of differentially and significantly expressed genes by E-EDGE

Differentially expressed genes out of global transcription data of 5584 genes were determined by using EDGE. The standard method yielded 1725 differentially expressed genes. Sampling time points were introduced as 0, 30, 60, 100, 140 and 180 minutes

corresponding to the real sampling times of 1-P dataset and time points 195, 210, 225 and 255 minutes were introduced corresponding to 0, 15, 30 and 60 minutes real sampling times within the 2-WT data set. A 15 minutes interval was artificially introduced between two data sets to connect these two separate data.

A total of 4344 genes that have at least 1.5 fold change between its maximum and minimum expression levels through the time scale were identified. In addition to that 3247 genes were determined to have standard deviations higher than 0.25.

A total of 4423 genes identified by these three approaches were then filtered as described in Section 3.3.6. in order to remove genes with low abundance. With this final step of E-EDGE 4141 genes were defined as differentially and significantly expressed and following analysis were applied on this subset of the transcriptome data.

#### 4.3.2. SOMs of differentially expressed genes identified by E-EDGE

Expression profiles of 4141 genes which were identified by E-EDGE were clustered by using SOMs and the number of clusters was intuitively determined as nine.

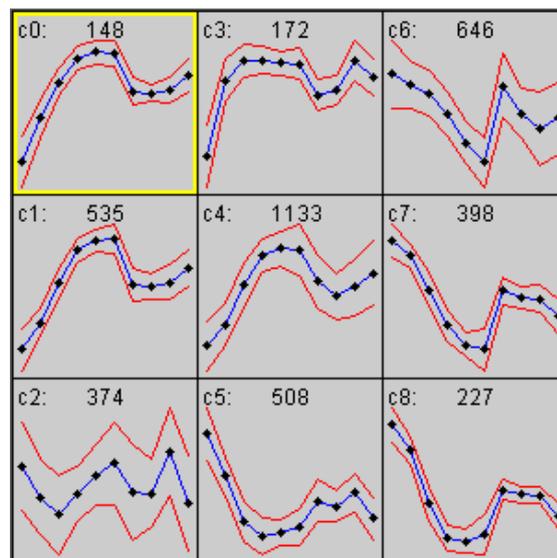


Figure 4.3. Self-organizing maps of expression profiles. The first six points and following four points represent the transcriptomic response to DNA damage introduced by MMS and UV respectively.

c0 and c1 clusters display an immediate up-regulation after exposure to MMS or UV (Figure 4.3). The number of genes and significantly associated biological processes and pathway are presented on Table 4.1.

Table 4.1. Selected GO biological process terms and pathways significantly associated with up-regulated clusters; c0 and c1. The number of the genes was given in parenthesis.

<b>c0 (148)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
vacuolar protein catabolic process	34	1.2E-28
cellular response to heat	34	5.1E-22
cellular carbohydrate catabolic process	14	1.0E-08
response to toxin	11	5.2E-08
alcohol catabolic process	11	1.6E-06
autophagy	13	8.4E-05
response to oxidative stress	10	8.7E-05
glycogen biosynthetic process	5	1.1E-04
sporulation resulting in formation of a cellular spore	15	4.1E-04
negative regulation of gluconeogenesis	4	8.8E-04
trehalose metabolic process	4	1.7E-03
glycoside metabolic process	4	2.2E-03
pentose catabolic process	3	4.9E-03
glucose catabolic process	6	5.3E-03
D-xylose metabolic process	3	7.3E-03
proteasomal ubiquitin-dependent protein catabolic process	7	9.4E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Starch and sucrose metabolism	10	2.2E-07
<b>c1 (535)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
vacuolar protein catabolic process	43	2.7E-16
cellular response to heat	49	1.3E-12
piecemeal microautophagy of nucleus	17	3.6E-09
CVT pathway	13	1.5E-08
cellular response to starvation	17	3.6E-07
glucose metabolic process	21	4.4E-04
sporulation resulting in formation of a cellular spore	36	4.7E-04
regulation of glucan biosynthetic process	6	5.2E-04
autophagic vacuole formation	6	5.2E-04
ubiquitin-dependent protein catabolic process	31	8.6E-04
glycogen metabolic process	10	8.8E-04
retrograde transport, endosome to Golgi	8	1.2E-03
trehalose metabolic process	6	1.6E-03
protein modification by small protein conjugation or removal	24	1.7E-03
glycoside metabolic process	6	2.6E-03
cellular aldehyde metabolic process	8	9.9E-03

Table 4.1. Selected GO biological process terms and pathways significantly associated with up-regulated clusters; c0 and c1. The number of the genes was given in parenthesis.

Cont.

Pathway	Count	P-Value
Regulation of autophagy	9	2.2E-05

c3 cluster displays an up-regulation until 60<sup>th</sup> minute of the MMS addition and then expression levels remain unchanged while after UV exposure up-regulation takes place until the 30<sup>th</sup> minutes followed with a slight down-regulation. Genes which compose the largest cluster, c4, were induced until 100<sup>th</sup> minute of MMS pulse and after 15<sup>th</sup> minute of the UV exposure (Figure 4.3). The significantly associated biological process terms and pathways are presented in Table 4.2.

Table 4.2. Selected GO biological process terms and pathways significantly associated with the clusters c3 and c4. The number of the genes was given in parenthesis.

<b>c3 (172)</b>		
GO Biological Process Term	Count	P-Value
arginine biosynthetic process	6	4.0E-06
response to inorganic substance	10	4.4E-06
iron-sulfur cluster assembly	7	1.4E-05
ornithine metabolic process	5	2.1E-05
response to toxin	9	3.8E-05
oxidation reduction	25	4.7E-05
response to reactive oxygen species	6	8.3E-05
response to drug	13	1.9E-04
cofactor biosynthetic process	11	8.6E-04
response to arsenic	4	3.2E-03
cellular response to heat	13	5.2E-03
sulfur metabolic process	9	7.3E-03
Pathway	Count	P-Value
Arginine and proline metabolism	5	1.7E-03
<b>c4 (1133)</b>		
GO Biological Process Term	Count	P-Value
ubiquitin-dependent protein catabolic process	70	8.3E-09
DNA repair	79	2.9E-06
protein complex assembly	76	8.2E-06
mitochondrion inheritance	15	1.0E-03
post-Golgi vesicle-mediated transport	27	1.2E-03
endosome transport	24	1.2E-03
transcription initiation from RNA polymerase II promoter	19	1.6E-03
protein modification by small protein conjugation	34	4.0E-03

Table 4.2. Selected GO biological process terms and pathways significantly associated with the clusters c3 and c4. The number of the genes was given in parenthesis. Cont.

membrane fusion	25	4.8E-03
vesicle docking during exocytosis	9	6.1E-03
actin cytoskeleton organization	32	8.9E-03
mRNA processing	55	8.9E-03
histone deacetylation	12	9.2E-03
chromatin silencing at telomere	21	9.9E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	31	2.6E-18

A total of 374 genes which compose cluster c2 display expression profiles with the highest standard deviation. MMS addition causes an upregulation after 60<sup>th</sup> minute following an initial down-regulation. UV exposure has the opposite effect on expression levels. The expression levels of the genes were induced between 30<sup>th</sup> to 60<sup>th</sup> minutes. The biological process GO terms and pathways associated with this cluster was given in Table 4.3.

Table 4.3. Selected GO biological process terms and pathways significantly associated with cluster c2. The number of the genes was given in parenthesis.

<b>c2 (374)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
translational elongation	39	5.2E-04
positive regulation of histone modification	4	2.4E-03
spliceosome assembly	6	9.3E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Spliceosome	10	3.1E-03

Down-regulation of 508 genes of c5 starts immediately after the pulse and stops at 100<sup>th</sup> minute of MMS addition while down-regulation starts at 60<sup>th</sup> minute after UV exposure. The biological process GO terms and pathways associated with this cluster was given in Table 4.4.

Table 4.4. Selected GO biological process terms and pathways significantly associated with the down-regulated cluster c5. The number of the genes was given in parenthesis.

<b>c5 (508)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
protein amino acid N-linked glycosylation	16	7.7E-06
polyadenylation-dependent ncRNA catabolic process	9	6.4E-05
transcription from RNA polymerase III promoter	12	6.6E-05
ncRNA catabolic process	10	1.1E-04
tRNA catabolic process	8	2.1E-04
RNA fragment catabolic process	6	4.7E-04
tRNA wobble uridine modification	9	1.2E-03
mitochondrial electron transport, cytochrome c to oxygen	6	2.4E-03
GPI anchor biosynthetic process	9	2.7E-03
exonucleolytic trimming to generate mature 3'-end of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	7	3.2E-03
U4 snRNA 3'-end processing	6	3.6E-03
maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	16	4.9E-03
mRNA catabolic process	16	6.9E-03
one-carbon metabolic process	15	9.7E-03
cell wall glycoprotein biosynthetic process	6	9.7E-03
mannoprotein biosynthetic process	6	9.7E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
High-mannose type N-glycan biosynthesis	8	4.5E-05
Pyrimidine metabolism	16	2.3E-03
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	8	4.1E-03
Purine metabolism	18	5.2E-03

MMS impulse causes repression of the genes which compose clusters c6 and c7 for the entire time interval of the experiment while it causes repression of the genes reside in cluster c8 until 100<sup>th</sup> minute. Exposure to UV also represses the genes in clusters c6, c7 and c8 in a way that the repression disappears at 30<sup>th</sup> minute for cluster c6 and accelerates for clusters c7 and c8. The biological process GO terms and pathways associated with these three cluster were given in Table 4.5.

Table 4.5. Selected GO biological process terms and pathways significantly associated with the down-regulated cluster c6, c7 and c8. The number of the genes was given in parenthesis

<b>c6</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
mitochondrial translation	44	4.5E-17
tRNA aminoacylation for protein translation	15	6.2E-05
RNA elongation from RNA polymerase II promoter	18	2.2E-04
nucleosome organization	17	1.3E-03
chromatin assembly or disassembly	17	2.7E-03
regulation of cell cycle	35	3.1E-03
cell division	59	3.7E-03
cytoskeleton organization	41	3.9E-03
intracellular protein transmembrane transport	21	4.3E-03
mitochondrial respiratory chain complex assembly	9	5.0E-03
conjugation with cellular fusion	25	6.0E-03
DNA packaging	14	7.8E-03
protein targeting to mitochondrion	14	7.8E-03
mitosis	33	8.7E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Aminoacyl-tRNA biosynthesis	15	3.0E-04
Ribosome	31	3.6E-04
<b>c7</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
rRNA processing	46	1.3E-09
ergosterol biosynthetic process	9	1.9E-04
ribosomal large subunit assembly	11	2.8E-04
ribosomal small subunit assembly	7	6.1E-04
regulation of translational fidelity	6	1.6E-03
cellular metabolic compound salvage	13	1.9E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Ribosome	58	1.6E-28
<b>c8</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ribosome biogenesis	110	8.5E-75
rRNA processing	93	3.6E-71

Table 4.5. Selected GO biological process terms and pathways significantly associated with the down-regulated cluster c6, c7 and c8. The number of the genes was given in parenthesis. Cont.

ncRNA processing	101	1.3E-65
rRNA modification	7	2.9E-05
positive regulation of transcription from RNA polymerase I promoter	5	1.2E-03
snRNA pseudouridine synthesis	4	1.3E-03
ribosomal large subunit export from nucleus	5	3.7E-03
cellular metabolic compound salvage	9	5.0E-03
tRNA methylation	5	5.9E-03
regulation of translation	16	8.3E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
RNA polymerase	10	2.1E-07
Pyrimidine metabolism	13	1.6E-06
Ribosome	14	1.3E-04
Purine metabolism	12	1.4E-04

#### 4.3.3. WGCNA of differentially expressed genes identified by E-EDGE

A co-expression network was constructed using differentially and significantly expressed genes identified by E-EDGE and analyzed by WGCNA. Power function was selected as an adjacency function and parameterization of the power function was identified according to fitness of the resulting co-expression network to scale free criteria. Scale free topology could be attained for  $\beta = 14$  (Figure 4.4).

Resulting adjacency matrix transformed into a topological overlap matrix (TOM) and then dissimilarities based on TOM were hierarchically clustered by using WGCNA R Package. Minimum module size was set to be 30 and module pairs with higher correlation than 0.90 between their representative expression profiles were merged. 20 modules were identified representing the response of the cells to DNA damage stress (Figure 4.5).

Hierarchical network and heat map of the modules based on correlations between their eigengene profiles provide clues about the organization of the modules. Modules are organized in five groups.

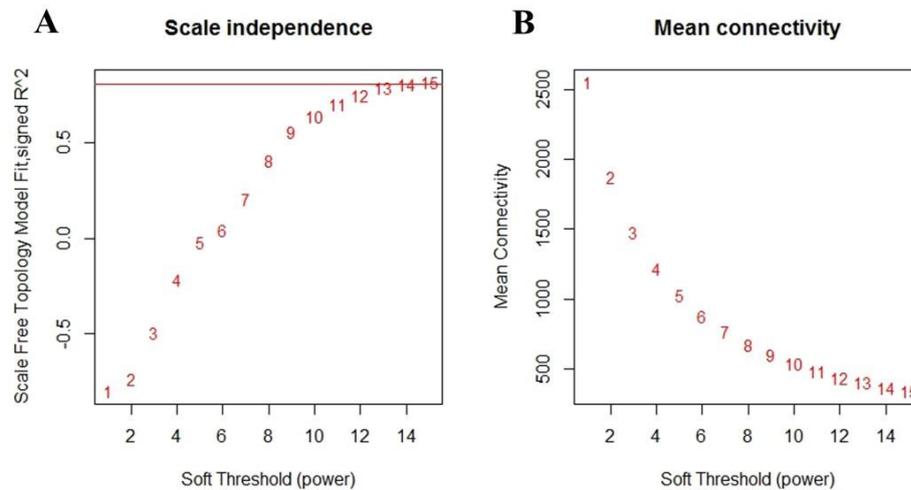


Figure 4.4. Fitness to scale-free nature (A) and mean connectivity (B) as a function of parameterization of the power function.

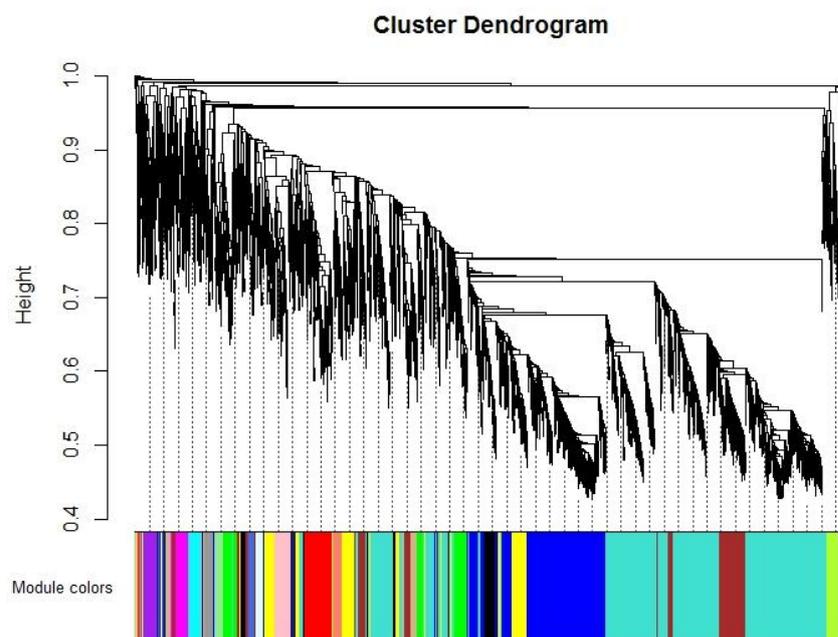


Figure 4.5. Hierarchical clustering dendrogram of genes identified by E-EDGE

The first group which has the closest correlation between their eigengenes consists of blue, pink, light-cyan and midnight-blue modules, the second group consists of green-yellow, green, light-green, brown and tan modules, the third group consists of cyan and grey60 modules, the fourth group consists of purple, magenta and red modules and turquoise, black, yellow, royal-blue and salmon modules compose the last group. Light yellow module

displays the most distinct eigengene profile which is not included within any of these five groups (Figure 4.6).

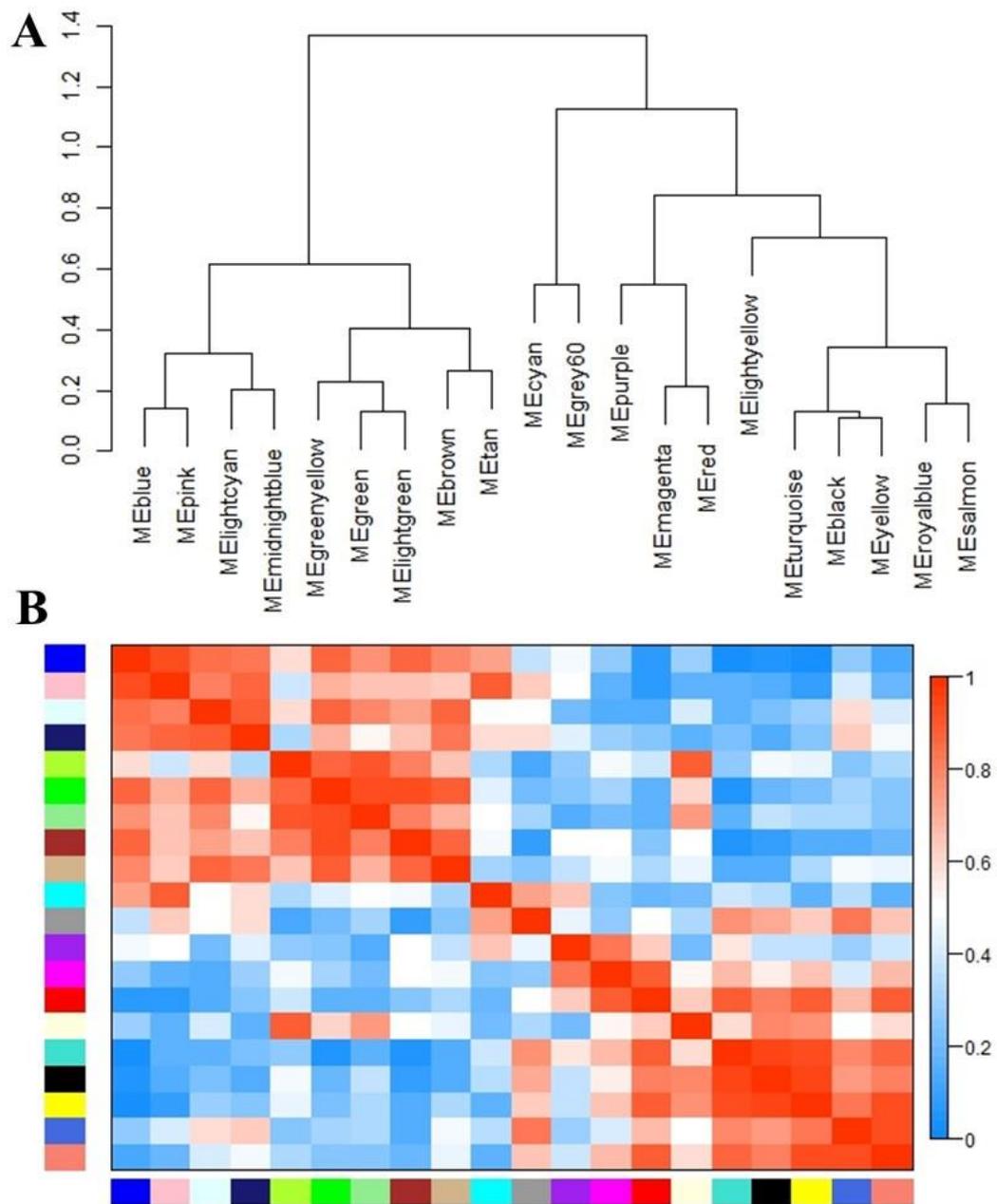


Figure 4.6. Meta-analysis of the modules identified by WGCNA for E-EDGE genes. Hierarchical organization of the modules (A) and heat-map of the modules (B).

By using the re-allocation process 3750 genes were redistributed to the 20 modules with a positive correlation ( $PCC > 0.75$ ) and remaining genes were flagged as outlier (Table

4.6). During the re-allocation process negatively correlated new modules were also constructed and because of the small sizes of these newly constructed modules they were also treated as outlier.

Table 4.6. Module ID's and sizes of E-EDGE genes determined by WGCNA after *signedKME* pipeline.

<u>Module</u>	<u>Module</u>	<u>Module</u>	<u>Module</u>	<u>Module</u>	<u>Module</u>	<u>Module</u>	<u>Module</u>
<u>ID</u>	<u>Size</u>	<u>ID</u>	<u>Size</u>	<u>ID</u>	<u>Size</u>	<u>ID</u>	<u>Size</u>
ME1	1437	ME6	170	ME11	78	ME16	55
ME2	599	ME7	126	ME12	76	ME17	55
ME3	375	ME8	99	ME13	71	ME18	52
ME4	348	ME9	86	ME14	71	ME19	48
ME5	194	ME10	83	ME15	68	ME20	45

Furthermore the average expression profiles of these modules were used to identify the time dependent up- or down regulated biological processes or pathways in response to DNA damage stress (Figure 4.7).

DNA damage stresses induce the genes in ME1 and ME6 modules. The induction in ME6 by MMS addition ends after one hour while the induction on account of UV exposure takes place between 15<sup>th</sup> and 30<sup>th</sup> minutes (Figure 4.7). The biological process GO terms and pathways associated with these two modules were provided in Table 4.7.

The genes involved in proteasomal ubiquitin-dependent protein catabolic process, sporulation resulting in formation of a cellular spore, CVT pathway and cellular response to starvation process were clustered in the largest module, ME1 and genes in ME6 module involved in response to toxin, oxidation reduction, response to hydrogen peroxide and cell redox homeostasis processes. Both of the modules were found to be significantly associated with cellular response to heat and vacuolar protein catabolic process biological process term. Proteasome and glutathione metabolism pathways were identified to be linked to ME1 and ME6 modules, respectively (Table 4.7).

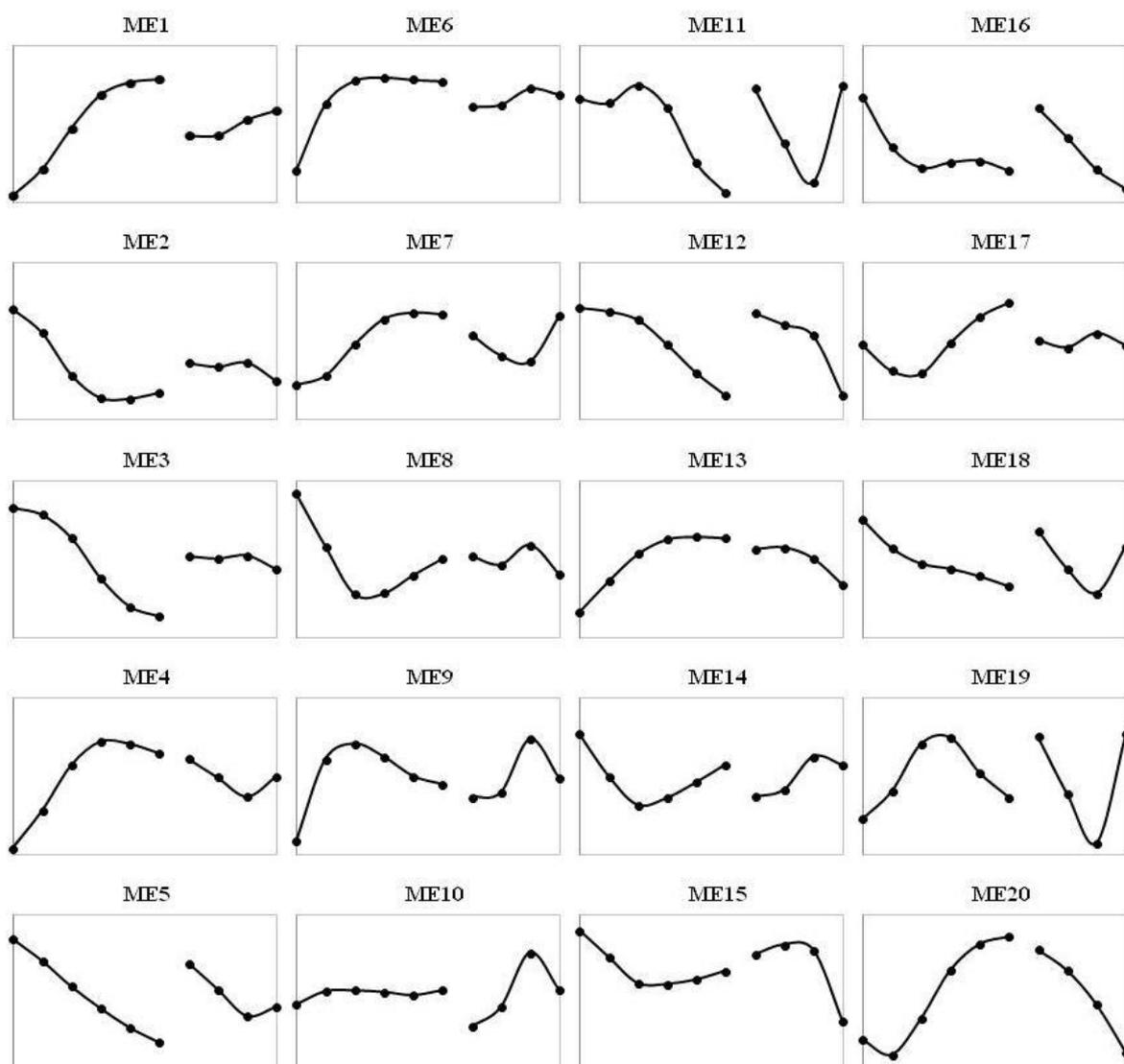


Figure 4.7. Average expression profiles of the positively correlated modules identified by WGCNA of the genes identified by E-EDGE (x-axis is sampling points and y-axis is fold change).

Table 4.7. Selected GO biological process terms and pathways significantly associated with ME1 and ME6 modules.

ME1		
GO Biological Process Term	Count	P-Value
vacuolar protein catabolic process	70	3.3E-26
cellular response to heat	72	4.8E-14
sporulation resulting in formation of a cellular spore	66	4.3E-07
proteasomal ubiquitin-dependent protein catabolic process	31	3.7E-06
late endosome to vacuole transport	18	1.5E-05
protein ubiquitination	31	1.5E-05

Table 4.7. Selected GO biological process terms and pathways significantly associated with ME1 and ME6 modules. Cont.

CVT pathway	12	8.6E-05
cellular response to starvation	18	2.3E-04
vesicle docking during exocytosis	10	3.1E-04
retrograde transport, endosome to Golgi	11	5.0E-04
D-xylose metabolic process	6	6.1E-04
membrane fusion	24	9.5E-04
proteasome assembly	10	9.7E-04
regulation of actin filament polymerization	7	1.0E-03
negative regulation of gluconeogenesis	7	1.0E-03
mRNA polyadenylation	10	1.6E-03
response to osmotic stress	30	2.0E-03
post-Golgi vesicle-mediated transport	23	2.8E-03
pentose catabolic process	5	3.1E-03
organelle inheritance	21	4.9E-03
glycogen biosynthetic process	7	7.2E-03
aldehyde catabolic process	5	8.1E-03
autophagic vacuole formation	6	8.2E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	30	6.3E-18
<b>ME6</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
response to toxin	15	8.2E-10
oxidation reduction	34	1.2E-06
cellular response to heat	22	4.7E-06
glutathione metabolic process	6	1.8E-04
response to hydrogen peroxide	4	2.7E-03
cell redox homeostasis	6	3.9E-03
glycerol ether metabolic process	4	4.0E-03
cofactor metabolic process	17	4.9E-03
vacuolar protein catabolic process	12	5.2E-03
lysine biosynthetic process	4	5.6E-03
iron-sulfur cluster assembly	5	6.4E-03
response to osmotic stress	11	6.4E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Glutathione metabolism	5	1.8E-03

The genes clustered in ME2 and ME3 were down-regulated upon MMS addition. They display a down-regulation with a 30 minutes delay after exposure to UV (Figure 4.7). Genes which compose ME2 and ME3 modules were identified to be associated with growth related processes. Purine and pyrimidine metabolism related pathways were found to be significantly linked to ME2 module and ribosome pathway was identified to be related to ME3 module (Table 4.8).

Table 4.8. Selected GO biological process terms and pathways significantly associated with clusters ME2 and ME3.

<b>ME2</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ncRNA metabolic process	195	5.3E-81
ribosome biogenesis	183	4.6E-80
ribosome export from nucleus	22	5.2E-09
transcription from RNA polymerase III promoter	19	6.8E-09
tRNA methylation	14	7.2E-09
cellular metabolic compound salvage	24	1.0E-06
regulation of translation	44	1.2E-05
one-carbon metabolic process	24	8.4E-05
cellular macromolecular complex assembly	60	3.3E-04
protein amino acid glycosylation	21	4.1E-04
positive regulation of transcription from RNA polymerase I promoter	7	1.7E-03
ribonucleoside monophosphate metabolic process	11	2.4E-03
snRNA pseudouridine synthesis	5	2.9E-03
exonucleolytic trimming during rRNA processing	8	5.8E-03
ergosterol biosynthetic process	9	6.5E-03
pyrimidine nucleotide metabolic process	7	6.7E-03
box C/D snoRNA metabolic process	6	7.1E-03
cell wall mannoprotein biosynthetic process	7	9.6E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
RNA polymerase	22	2.1E-13
Purine metabolism	36	1.2E-10
Pyrimidine metabolism	30	9.2E-10
High-mannose type N-glycan biosynthesis	8	2.8E-04
<b>ME3</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
translation	140	2.6E-32
ribosome biogenesis	54	7.7E-07
ergosterol biosynthetic process	11	4.8E-06
regulation of translational fidelity	8	1.7E-05
rRNA export from nucleus	13	4.8E-05
cellular metabolic compound salvage	16	7.2E-05
tRNA aminoacylation for protein translation	11	4.3E-04
fatty acid elongation	4	1.7E-03
posttranslational protein targeting to membrane, translocation	5	1.9E-03
protein amino acid N-linked glycosylation	10	9.8E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Ribosome	82	2.1E-47

Average expression profiles of ME4 and ME7 modules display an increase within the first hour of MMS addition and a decrease within the first half hour of the UV exposure (Figure 4.7).

Functional annotations of the genes which compose ME4 module are similar to the functional annotations determined for ME1 module while ME7 was found to be related to DNA repair and processes linked to cell cycle (Table 4.7; Table 4.9).

Table 4.9. Selected GO biological process terms and pathways significantly associated with ME4 and ME7 modules.

<b>ME4</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
vacuolar protein catabolic process	19	7.0E-07
cellular response to heat	17	2.3E-03
glycogen metabolic process	6	6.3E-03
membrane organization	20	9.2E-03
response to inorganic substance	7	9.6E-03
<b>ME7</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
DNA repair	28	6.7E-07
protein refolding	7	2.7E-05
glutamate biosynthetic process	5	1.2E-03
meiosis I	11	1.3E-03
cellular response to heat	17	1.6E-03
M phase of meiotic cell cycle	16	4.8E-03
DNA catabolic process, endonucleolytic	5	7.8E-03
DNA integrity checkpoint	5	9.3E-03
coenzyme catabolic process	6	9.3E-03
mitotic sister chromatid segregation	9	9.4E-03
mitochondrion organization	20	9.6E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Meiosis	13	3.5E-03
Homologous recombination	5	6.2E-03

Down-regulated modules ME5 and ME18 upon MMS addition display a down-regulation until 30<sup>th</sup> minute of exposure to UV where expression levels start to slightly recover (Figure 4.7).

Both of the modules have genes which are significantly enriched in cell cycle related biological processes like regulation of cell cycle and mitosis. The genes which compose ME5 module was also determined to be associated with cell cycle pathway (Table 4.10)

Table 4.10. Selected GO biological process terms and pathways significantly associated with ME5 and ME18 modules.

<b>ME5</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
regulation of transcription from RNA polymerase II promoter	19	2.1E-04
cytoskeleton organization	17	2.5E-04
cell cycle phase	25	2.9E-04
regulation of cell cycle	13	2.8E-03
chromatin modification	13	4.2E-03
mitosis	12	7.4E-03
cellular bud site selection	7	8.2E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Cell cycle	15	1.6E-06
<b>ME18</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
regulation of cell cycle	11	2.3E-05
cytoskeleton-dependent intracellular transport	4	2.1E-03
negative regulation of DNA metabolic process	4	2.1E-03
sister chromatid segregation	6	2.2E-03
mitosis	8	3.3E-03
M phase of mitotic cell cycle	8	3.6E-03
meiotic cell cycle	8	5.1E-03
regulation of DNA replication	4	5.4E-03
establishment of mitotic spindle orientation	3	6.5E-03

Average expression profiles of ME8 and ME9 modules show opposite trends upon MMS addition while they behave similarly after UV exposure (Figure 4.7).

Oxidative phosphorylation pathway related genes in ME8 module were found to be repressed after MMS addition for an hour and then upregulated while arginine and proline metabolism pathway related genes in ME9 were initially induced and then repressed after 60<sup>th</sup> minute. Both pathways display similar expression profiles after UV exposure as they upregulated after 15<sup>th</sup> minute and down-regulated after 30<sup>th</sup> minute (Table 4.11).

Table 4.11. Selected GO biological process terms and pathways significantly associated with ME8 and ME9 modules.

<b>ME8</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
rRNA processing	17	1.3E-05
nucleobase, nucleoside and nucleotide biosynthetic process	11	2.2E-04
tRNA modification	7	2.1E-03
purine nucleotide biosynthetic process	7	3.4E-03
peptidyl-diphthamide metabolic process	3	3.7E-03
oxidative phosphorylation	6	5.8E-03
purine salvage	3	7.6E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Oxidative phosphorylation	6	9.9E-03
<b>ME9</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
arginine biosynthetic process	7	1.8E-11
amine biosynthetic process	12	2.0E-09
response to drug	7	2.1E-04
ornithine biosynthetic process	3	3.0E-04
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Arginine and proline metabolism	5	8.4E-05

Genes in ME11 module were down-regulated with a 100 minutes delay after MMS addition while they immediately down-regulated until 30<sup>th</sup> minute of UV exposure (Figure 4.7). ME11 module was identified to be significantly associated with mitochondrial translation, regulation of exit from mitosis and mitochondrial respiratory chain complex IV assembly biological processes (Table 4.12).

Table 4.12. Significantly enriched GO biological process terms of ME11 module.

<b>ME11</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
mitochondrial translation	25	1.3E-27
mitochondrion organization	30	9.8E-22
translation	34	1.5E-14
regulation of exit from mitosis	5	2.6E-04
mitochondrial respiratory chain complex IV assembly	3	8.5E-03

Average expression profile of the genes which compose ME12 module displays a continuous down-regulation which accelerates after 60<sup>th</sup> and 30<sup>th</sup> minutes of MMS addition and UV exposure, respectively (Figure 4.7). The genes residing in ME12 are enriched in nucleosome assembly, cytokinesis and cell separation during cytokinesis biological processes (Table 4.13).

Table 4.13. Significantly enriched GO biological process terms of ME12 module.

<b>ME12</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
nucleosome organization	7	1.2E-04
chromatin assembly or disassembly	7	1.7E-04
cytokinesis, completion of separation	4	3.3E-04
nucleosome assembly	5	3.8E-04
cell division	14	4.6E-04
cytokinesis	8	6.3E-04
chromatin assembly	5	8.3E-04
cell separation during cytokinesis	4	8.8E-04
DNA packaging	5	4.9E-03

The genes in ME13 module which responses MMS addition by induction and responses UV exposure by repression were found to be significantly enriched in only three biological process terms; regulation of transcription, regulation of DNA-dependent transcription and regulation of RNA metabolic process (Figure 4.7).

ME14 module consists of genes related to spliceosome pathway and nuclear mRNA splicing via spliceosome process (Table 4.14). Expression profiles of these genes display a repression until 60<sup>th</sup> minute of MMS addition and display an induction until 30<sup>th</sup> minute of UV exposure (Figure 4.7).

Table 4.14. Significantly enriched GO biological process terms of ME14 module.

<b>ME14</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
mRNA metabolic process	8	2.6E-03
RNA splicing	6	3.1E-03
nuclear mRNA splicing, via spliceosome	5	5.4E-03
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	5	5.8E-03
RNA splicing, via transesterification reactions	5	7.3E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Spliceosome	4	9.8E-03

Cell cycle related genes which reside in ME16 were found to be down-regulated continuously upon UV exposure and down-regulated within the first half hour after MMS addition (Figure 4.7; Table 4.15).

Table 4.15. Selected GO biological process terms and pathways significantly associated with ME16 modules.

<b>ME16</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
cell cycle process	17	9.8E-05
nucleosome assembly	5	1.8E-04
mitotic cell cycle	12	3.9E-04
M phase	12	7.7E-04
positive regulation of nucleobase-containing compound metabolic process	8	3.3E-03
positive regulation of macromolecule metabolic process	8	7.5E-03
negative regulation of transcription from RNA polymerase II promoter	5	9.0E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Cell cycle	6	4.8E-03

The genes which compose ME17 module were irresponsive to UV while they were repressed initially until 30<sup>th</sup> minute of MMS addition whereafter expression levels were increased. These genes do not have any significant biological process terms association while the only pathway statistically enriched with the genes of the module was found to be one-carbon-pool by folate.

UV causes repression of genes which compose ME19 module for half an hour then transcription of the genes recovers while MMS addition leads to a slight increase and a recovery after 100 minutes (Figure 4.7). The module was determined to be associated with mitochondrion organization and translation processes and Aminoacyl-tRNA biosynthesis pathway (Table 4.16).

Average expression profile of the last module, ME20 increases with a 30 minute delay after MMS addition and decreases continuously after exposure to UV (Figure 4.7). The genes were found to be associated with four biological process terms; regulation of DNA-dependent transcription, regulation of RNA metabolic process, regulation of transcription and transcription.

Table 4.16. Significantly enriched GO biological process terms of ME19 module.

<b>ME19</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
mitochondrial translation	14	2.5E-12
mitochondrion organization	20	4.2E-12
translation	21	3.0E-06
tRNA aminoacylation	6	5.4E-05
amino acid activation	6	5.4E-05
tRNA aminoacylation for protein translation	5	5.2E-04
tRNA aminoacylation for mitochondrial protein translation	3	6.1E-03
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Aminoacyl-tRNA biosynthesis	6	1.9E-05

ME10 and ME15 were found to be not associated with any process or pathways.

#### 4.3.4. Regulation of the response to DNA damage

Expression profiles of 4003 genes out of 4141 genes identified by E-EDGE with known TF interactions were analyzed by DREM. Analysis of this subset by DREM revealed that DNA damage triggers complex regulatory events where 22 transcription factors (TFs) drives transcriptomic response into 9 branches associated with 162 to 795 genes at the end of time course (Figure 4.8).

Transcriptional profiles of the 4003 genes immediately splits into two branches. A set of eight regulators (Yap7p, Msn2p, Rpn4p, Cad1p, Yap1p, Msn4p, Aft2p and Hsf1p) which are mostly related to stress response mediate the regulation of the up-regulated branch of genes. These genes bifurcates at the 30<sup>th</sup> minute of MMS addition where continuously induced sub-branch of genes were determined to be controlled by Msn2p, Msn4p, Yap7p, Skn7p and Ume6p. Nuclear response regulator Skn7p drives the further induction of the genes which bifurcate again at the consecutive time point (Figure 4.8).

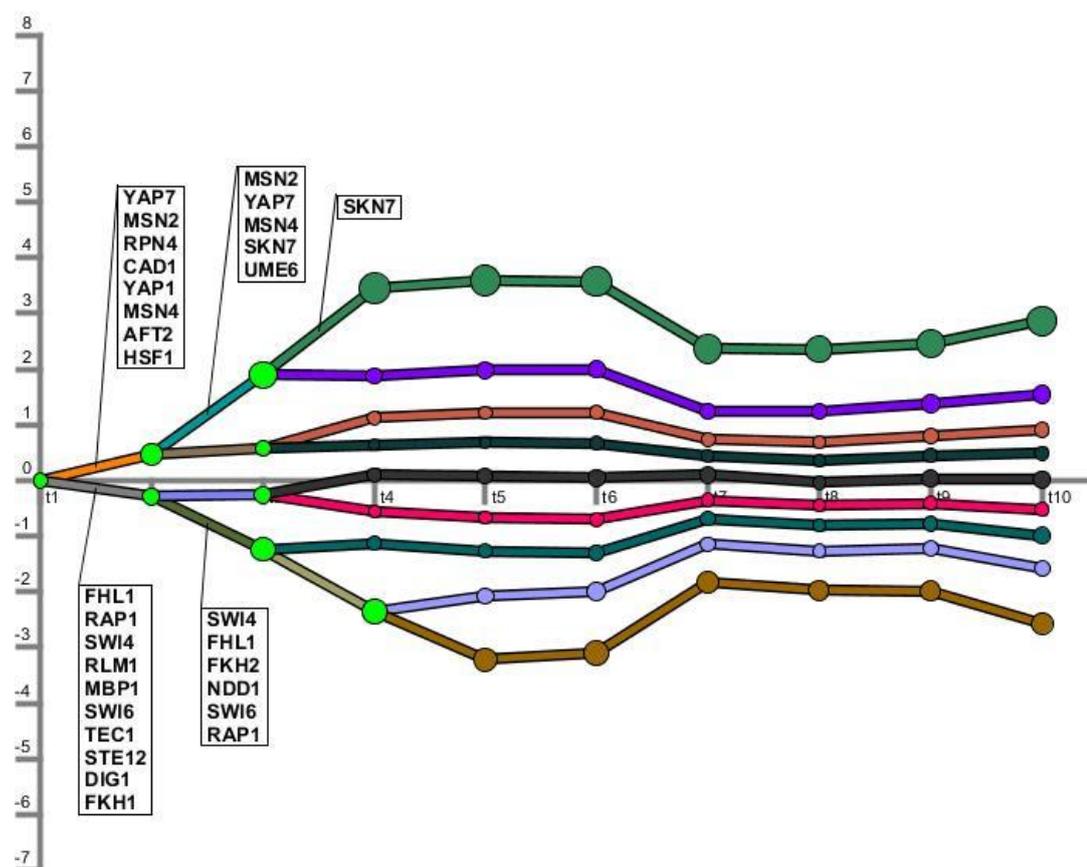


Figure 4.8. Dynamic regulatory map of differentially expressed genes. X-axis was arranged as uniform sample points. The first six points and following four points represent the transcriptomic response to DNA damage introduced by MMS and UV respectively.

Another 10 TFs (Fhl1p, Rap1p, Swi4p, Rlm1p, Mbp1p, Swi6p, Tec1p, Ste12p, Dig1p and Fkh1p) which control cell cycle, mating and ribosomal proteins were identified to be master regulators of the down-regulated genes at the beginning. Ndd1p which is a transcriptional activator essential for nuclear division, Fkh2p which is the paralog of Fkh1p and four other TFs (Swi4p, Fhl1p, Swi6p, Rap1p) among these 10 TFs continue to play role in repression of a sub-branch of genes at the following split (Figure 4.8).

#### 4.3.5. NP analysis of the DNA damage

A protein-protein interaction network was constructed for NP analysis, consisting of physical interactions verified by more than one different experimental method or mentioned in more than one articles in MIPS, MINT, DIP and BIOGRID databases (Mewes *et al.*, 2002; Chatr-Aryamontri *et al.*, 2007; Xenarios *et al.*, 2000; Stark *et al.*, 2006). There were 5,985

such interactions among the 5576 proteins whose corresponding expression data is available. PCC cut-off was selected to be more stringent as 0.7 and -0.7 for correlation and anti-correlation in NP analysis. NP networks for MMS addition (1-P) and UV exposure (2-WT) were constructed separately using the reference ppi network of 5576 nodes and 5985 edges. Topological properties of the resulting NP networks are presented in Table 4.17.

Table 4.17. Topological properties of NP networks.

	1-P	2-WT
Number of correlated edges	3025	2289
Number of anti-correlated edges	1021	548
Number of total edges	4046	2837
Number of nodes	1789	1545

NP networks were manually dissected into two anti-correlated clusters (R and B) so that each cluster has less than 1% intra-cluster anti-correlated interactions. Expression profiles of each cluster in time for each data set were plotted (Figure 4.9.).

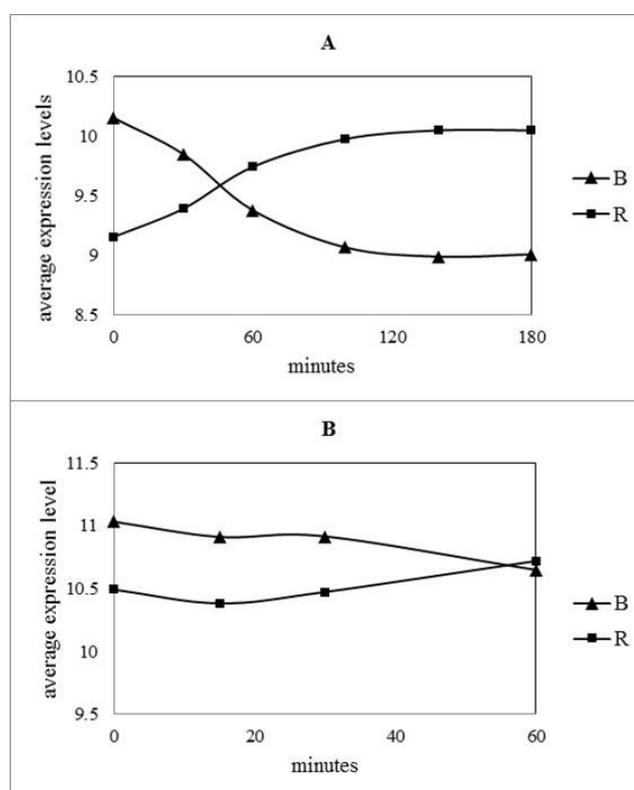


Figure 4.9. Average expression levels of R and B modules in; (A) 1-P and (B) 2-WT.

In response to MMS addition 969 genes clustered in an upregulated module (R) out of physically interacting NP network while expression levels of 820 genes (B) were found to

be repressed (Figure 4.9a). The biological process GO terms and pathways significantly associated with the upregulated (R) module is given in Table 4.18.

Table 4.18. Selected GO biological process terms and pathways significantly associated with R module in response to MMS.

<b>R module (1-P)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ubiquitin-dependent protein catabolic process	94	2.0E-22
protein complex assembly	96	5.9E-14
cell cycle process	166	1.8E-12
protein ubiquitination	46	2.3E-12
RNA biosynthetic process	88	1.1E-11
piecemeal microautophagy of nucleus	25	3.2E-11
sporulation resulting in formation of a cellular spore	84	4.0E-11
post-Golgi vesicle-mediated transport	38	1.8E-09
DNA repair	87	2.0E-09
actin filament organization	33	2.3E-08
histone deacetylation	19	8.8E-08
mRNA metabolic process	88	1.1E-07
vesicle fusion	22	1.1E-07
vesicle docking during exocytosis	14	1.4E-07
protein amino acid phosphorylation	52	5.7E-07
protein targeting to vacuole	31	6.0E-07
cell division	104	6.8E-07
response to nutrient levels	31	9.1E-07
endosome transport	30	1.0E-06
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	34	3.4E-21
Regulation of autophagy	16	4.3E-09
Ubiquitin mediated proteolysis	26	2.2E-07
Cell cycle	49	2.0E-06
Basal transcription factors	15	3.0E-05
Homologous recombination	13	7.0E-05
Meiosis	44	1.7E-04
Endocytosis	16	1.2E-03
Nucleotide excision repair	16	2.5E-03

The biological process GO terms and pathways significantly associated with the down-regulated module B in response to MMS were tabulated in Table 4.19.

Table 4.19. Selected GO biological process terms and pathways significantly associated with B module in response to MMS.

<b>B module (1-P)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ribosome biogenesis	184	3.4E-58
rRNA processing	142	4.8E-53
ncRNA processing	170	9.1E-51
RNA modification	48	1.3E-09
cell division	100	4.9E-09
intracellular protein transmembrane transport	39	8.4E-09
polyadenylation-dependent ncRNA catabolic process	14	2.5E-07
tRNA catabolic process	13	3.8E-07
ribosome export from nucleus	22	8.7E-07
protein import into nucleus	26	1.0E-06
nucleus organization	30	1.3E-06
regulation of cyclin-dependent protein kinase activity	14	1.6E-06
RNA transport	40	1.6E-06
cell cycle phase	107	1.9E-06
U4 snRNA 3'-end processing	11	2.7E-06
tRNA processing	38	3.7E-06
regulation of translation	55	3.7E-06
mitosis	54	4.1E-06
nuclear mRNA splicing, via spliceosome	36	4.9E-06
chromosome segregation	51	6.3E-06
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	17	2.0E-05
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
RNA polymerase	27	7.8E-17
Pyrimidine metabolism	39	3.3E-12
Purine metabolism	43	1.7E-10
RNA degradation	32	1.8E-10
Cell cycle	49	4.0E-08
Spliceosome	29	8.3E-07

A total of 694 genes form an up-regulated module in response to UV exposure (Figure 4.9b). Exposure to UV induces similar pathways and biological processes but its effect on catabolic processes seems to be less severe when compared to MMS addition (Table 4.18; Table 4.20).

Table 4.20. Selected GO biological process terms and pathways significantly associated with R module in response to UV exposure.

<b>R module (2-WT)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
DNA replication	58	2.1E-14
mitochondrial translation	36	4.8E-09
DNA repair	69	5.0E-09
microtubule cytoskeleton organization	34	1.4E-06
protein ubiquitination	30	1.5E-06
reproduction of a single-celled organism	57	3.2E-06
M phase of mitotic cell cycle	48	4.8E-06
protein transport	105	5.2E-06
actin filament organization	24	6.8E-06
mitotic recombination	17	7.6E-06
vesicle-mediated transport	83	8.0E-06
regulation of DNA metabolic process	23	1.0E-05
vesicle organization	23	1.4E-05
proteasomal protein catabolic process	27	1.4E-05
sister chromatid cohesion	17	2.6E-05
mismatch repair	14	3.1E-05
telomere maintenance	22	3.6E-05
cyclin catabolic process	9	3.9E-05
membrane organization	61	4.1E-05
cellular protein complex assembly	35	4.4E-05
cell cycle checkpoint	22	6.3E-05
regulation of protein complex assembly	11	6.9E-05
mitotic metaphase/anaphase transition	9	8.8E-05
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
Proteasome	34	1.5E-24
Cell cycle	50	2.2E-10
Meiosis	44	3.4E-07
DNA replication	17	3.8E-06
Mismatch repair	13	6.3E-06
Nucleotide excision repair	17	4.8E-05
Homologous recombination	12	5.1E-05
Ubiquitin mediated proteolysis	19	1.4E-04

A total of 870 genes compose a down-regulated module in response to UV exposure

(Figure 4.9b). Exposure to UV induces the same set of pathways and similar biological processes related to growth as in the case of MMS addition (Table 4.19; Table 4.21).

Table 4.21. Selected GO biological process terms and pathways significantly associated with B module in response to UV exposure.

<b>B module (2-WT)</b>		
<b>GO Biological Process Term</b>	<b>Count</b>	<b>P-Value</b>
ribosome biogenesis	175	1.1E-46
ncRNA metabolic process	167	3.3E-33
transcription, DNA-dependent	112	1.7E-28
chromatin organization	91	8.9E-17
nuclear export	53	1.3E-10
histone modification	45	2.5E-10
protein import into nucleus	31	1.8E-09
nuclear import	31	1.8E-09
RNA transport	46	1.2E-08
ribosome export from nucleus	25	1.7E-08
ribosome assembly	32	1.2E-07
ATP-dependent chromatin remodeling	20	2.8E-07
intracellular protein transmembrane transport	37	5.6E-07
cell division	97	1.1E-06
interphase of mitotic cell cycle	40	1.3E-06
RNA modification	43	2.4E-06
nuclear pore organization	19	3.0E-06
rRNA export from nucleus	21	7.2E-06
protein export from nucleus	19	8.7E-06
cytoskeleton organization	63	6.1E-05
G2/M transition of mitotic cell cycle	18	6.2E-05
cellular protein complex disassembly	24	8.3E-05
<b>Pathway</b>	<b>Count</b>	<b>P-Value</b>
RNA polymerase	22	5.1E-11
Spliceosome	32	1.2E-09
RNA degradation	29	5.0E-09
Purine metabolism	35	8.7E-07
Pyrimidine metabolism	28	7.7E-06
Cell cycle	39	1.0E-04

#### 4.4. Discussion

The clustering and analysis of response to DNA damage indicated that the genes involved in vacuolar and proteasomal ubiquitin-dependent protein catabolic processes were found to be induced after exposure to DNA damaging effectors. WGCNA and NP analysis of the same data sets provided further supportive information. These results are in good correlation with the fact that protein catabolism was reported to be induced by MMS addition (Shalem *et al.*, 2008).

Rpn4p which stimulates expression of proteasome genes was also identified as one of the master regulators of up-regulated genes by DREM. *RPN4* deletion strain was found to be MMS sensitive (Jelinsky *et al.*, 2000) and Rpn4p was also identified as one of the DNA damage responsive TFs (Workman *et al.*, 2006).

One set of genes involved in cellular response to heat was also found to be induced in response to DNA damage. WGCNA also identified a set of genes with the same association. However another smaller set of genes identified by WGCNA which is also enriched in cellular response to heat displays up-regulation until 100<sup>th</sup> minute after MMS addition while exposure to UV cause down-regulation for the first half hour before induction of the genes associated with the same biological process.

Heat shock protein Hsf1p was found to be a significant mediator of the up-regulation by DREM in correlation with up-regulation of cellular response to heat. It was reported that heat shock related genes including *HSF1* significantly response to DNA damage caused by MMS (Workman *et al.*, 2006).

The response to drug biological process term related *YAPI*, *ATR1*, *SNG1* and *ROD1* were identified to be immediately up-regulated within the first half hour of the exposure to both MMS and UV by SOMs and WGCNA. Yap1p was already reported to be differentially expressed after exposure to MMS (Workman *et al.*, 2006). Transcription factor was also identified by DREM as one of the key regulators of the up-regulated genes. There are some other genes related to response to drug biological process (*FLR1* and *CIN5*) whose expression levels were up-regulated after MMS addition but unchanged after UV irradiation.

*FLRI* is specifically related to drug transmembrane transport that might be the reason for its indifference to UV.

The same clusters identified by SOMs and WGCNA which were enriched in response to drug were also found to be enriched in arginine biosynthesis related biological processes and pathways. Individual investigation of arginine biosynthesis related genes in these clusters (*ARG1*, *ARG2*, *ARG3*, *ARG4*, *ARG5,6*, *ARG7* and *ARG11*) showed that MMS induces transcription of these genes within the first half hour while UV irradiation dependent upregulation starts after 15 minute which is followed by recovery after 30<sup>th</sup> minute. It was reported that *Candida albicans* upregulates its arginine biosynthesis in response to ROS (Jimenez-Lopez *et al.*, 2012) and it was also reported that arginine biosynthetic pathways are involved in human carcinogenesis (Lind, 2004). However the role of arginine biosynthetic pathway in DNA damage response is poorly understood so far.

Another set of genes whose expression levels were also observed to be up-regulated after MMS addition were found to be related to DNA repair process. Up-regulation of this group after UV irradiation starts with a 15 minutes delay. NP analysis verifies exactly the same observations while WGCNA supports these observations with a slight difference that a smaller set of genes enriched in DNA repair process was induced with a 30 minutes of delay after exposure to UV.

Nucleotide excision repair (NER) and mismatch repair (MMR) mechanisms were found to be activated DNA repair mechanisms in response to DNA damage. Interestingly base-pair excision repair (BER) mechanism could not be identified as responsive to DNA damage.

DNA damage signaling triggers Chk kinases and Chk kinases, in turn, activates DNA repair machinery (Friedberg *et al.*, 2005). The genes encoding two Chk kinases, *RAD53* and *DUN1*, which were identified in these sets are known to be involved in DNA damage process.

DNA packaging, nucleosome organization and chromatin assembly biological processes were identified to be down regulated upon exposure to both MMS and UV by all of the three approaches. Both Rap1p and Fkh2p can repress chromatin silencing and the both

were identified to be master regulators of down regulated genes by DREM. All these findings point out that cells try to unpack DNA in order to help DNA repair processes in response to DNA damage. It is commonly accepted that DNA processing events including DNA repair are inhibited by DNA packaging (Sancar *et al.*, 2004).

SOMs, WGCNA and NP analysis coordinately revealed that mitosis related genes were down-regulated upon exposure to MMS or UV. NP analysis of transcriptional response to MMS and UV also capture down-regulation of G2/M transition of mitotic cell cycle.

Gasch and colleagues reported that MMS and irradiation arrests cell cycle at different stages; MMS leads to S-phase arrest while UV cause an arrest at G2/M transition in correlation with the observed distinction (Gasch *et al.*, 2001).

The cell cycle regulators Mbp1p, Swi6p, Ndd1p, Fkh1p and Fkh2p were identified as master regulators of down-regulated genes by DREM.

Ribosome biogenesis and accompanying growth related processes were identified to be down-regulated after exposure to both MMS and UV by all of the three approaches; SOMs, WGCNA and NP analysis. Fhl1p which is a regulator of ribosomal protein transcription was also identified as one of the master controllers of down-regulated genes in the present study.

Jelinsky and Samson reported 13 ribosomal proteins whose transcripts were repressed more than three-fold by MMS treatment and concluded that alkylation exposure leads to a slowdown of global protein production (Jelinsky and Samson, 1999). Another study supports also the repression of ribosome biogenesis in DNA damaged cells through irradiation (Wade *et al.*, 2009).

Skn7p which is a nuclear response regulator and transcription factor was reported to be one of the central TFs in response to oxidative stress (Kelley and Ideker, 2009) and deletion of *SKN7* was reported to increase sensitivity to MMS (Brown *et al.*, 2006). Skn7p was identified by DREM analysis as one of the significant TFs while it could not be captured by Workman and colleagues (Workman *et al.*, 2006).

Autophagy related biological processes and regulation of autophagy pathway were identified to be upregulated by clustering of expression profiles. It was reported that DNA damage induces autophagy (Kang *et al.*, 2009; Rodriguez-Rocha *et al.*, 2011) and induced autophagy delays apoptosis (Abedin *et al.*, 2007) in human.

A striking difference between the two responses was observed to be the magnitude of the responses. Number of genes whose expression level changes more than two-fold was found to be 1989 and 216 after exposure to the MMS and UV, respectively. The difference supports that the magnitude of the stress caused by MMS is much more than UV-irradiation. In accordance with this conclusion Benton and colleagues have reported that transcriptional response of yeast cells to MMS and ionizing radiation is dose-dependent (Benton *et al.*, 2006).

Processes and pathways which are responsive to DNA damage like DNA repair, catabolism, ribosome biogenesis, cell cycle, response to heat, autophagy or DNA packaging were found to be affected by MMS and UV-irradiation in a similar fashion. Meiosis was determined to be upregulated only after exposure to MMS by WGCNA however NP analysis revealed that the process was up-regulated also by UV irradiation. Supporting results with different methods as implemented in the presents study provides more reliable outcomes.

#### **4.5. Future prospects**

The magnitude of the transcriptional responses could be synchronised by repeating the chemostat experiments with different doses of DNA damaging agents. As a future perspective, experimenting transcriptional response of yeast cells to different doses of doxorubicin and to different doses of DNA damaging agents under the same conditions following by a comprehensive and reliable data analyzing step as proposed and implemented in this study would bring a clearer understanding of underlying mechanisms of response to doxorubicin and DNA damage.

Activation of basepair excision repair process upon doxorubicin pulse and irresponsiveness of the process to MMS and UV should be investigated further. Especially the role of *OGGI* in this difference should be focused.

Topological analysis of the weighted co-expression network which was constructed by WGCNA can be investigated further by using WSL-EC. Topological centrality in co-expression networks and its biological implications may unveil additional information.

## 5. CONCLUSION

In the first part of the thesis (chapter 2), a novel global metric of centrality, weighted sum of loads eigenvector centrality (WSLEC), counting all eigenvectors was proposed. The performance of WSL-EC in the identification of topologically more important nodes that contribute to the integrity of a network and in capturing essential or biologically central nodes was tested in three biological networks and compared with the performances of four other commonly used metrics of centrality, DC, BC, EC and SC.

Hubs with off-the-scale connectivity (super-hubs) create a strong bias in topological centrality for DC, BC, EC and SC, whereas WSL-EC does not seem to be affected by the presence of super-hubs.

WSL-EC was found to outperform in capturing biologically central nodes, such as pathogen-interacting, HIV-1, cancer, ageing, and disease-related genes and genes, involved in immune system process and related to autoimmune diseases in the human interactome compared with DC, BC, EC or SC. The choice of metric of centrality is crucial, as different metrics focus on different topologies and these topological differences correspond to different biological roles.

In the following part of the study (chapter 3) dynamic transcriptional response induced by doxorubicin were measured and analyzed. Performances of different methodologies were investigated and several modifications were proposed wherever needed.

The clustering and analysis of response to doxorubicin indicated that the genes involved in DNA replication, mismatched repair, cell cycle and base excision repair pathways were found to remain without any change in their expression levels within the first five minutes after the induction of doxorubicin and they were down-regulated until the 60<sup>th</sup> minute where they start to be up-regulated. Another set of genes involved in proteasome also behaved similarly. However the expression levels of the set remained unchanged for 60 minutes and then up-regulated.

One set of genes whose expression levels were observed to be down regulated after 30 minutes consist of genes involved in meiosis, regulation of cell cycle, cell wall organization, response to osmotic stress and chromosome organization.

Induction of ribosome biogenesis and coherent down-regulation of vacuolar protein catabolic process were found to be overrepresented in upregulated and downregulated clusters, respectively. Up-regulation and down-regulation regarding these terms took place immediately after doxorubicin addition to the chemostat culture. All of the applied methods revealed similar findings which were in correlation with these observations.

Analysis of the expression profiles indicated that the expression levels of the genes involved in siderophore transport remained unchanged until one hour before sharply up regulated and the genes involved in cellular iron ion homeostasis were found to be up-regulated after 5<sup>th</sup> minute.

In this chapter an extended version of EDGE was also proposed in order to identify differentially expressed genes out of a dynamic dataset. Clustering approaches were applied to this subset assuming that biologically irrelevant or inactive data was eliminated which can improve biological associations of the resulting modules or clusters. Identifying differentially expressed genes by E-EDGE and then implementing clustering analysis on the resulting significant subset of the data can strengthen the results and unveil additional biological data.

Although size of the subset identified by E-EDGE is more than twice the size of the subset identified by variance filtering in SOMs analysis, biological significance of the resulting clusters were improved. SOMs of the genes identified by E-EDGE yielded nine clusters with distinct biological roles.

Analysis of the genes identified by E-EDGE revealed some of the stress responsive TFs (Msn2p, Msn4p and Mcm1p) and regulators of the processes like proteasome (Rpn4p), glucose metabolism (Sip4p and Nrg1p), copper-sensing (Mac1p) and ribosome biogenesis (Fhl1p) which could not be identified when the global transcriptome was used.

A reallocation pipeline was also developed and proposed to improve WGCNA. The reallocation process yielded stronger biological attachments for the modified modules.

Under the light of these results it can be proposed to track the following pipeline for dynamic transcriptome data analysis (Figure 4.2);

- Identify differentially expressed genes by E-EDGE which is an extended version of EDGE approach,
- Determine modules which represent the dataset by WGCNA,
- Re-allocate genes to the modules identified by WGCNA by emphasizing positive correlations,
- Crosscheck the modules by using SOMs,
- Integrate additional omics data if it is needed. DREM can be used to integrate regulome and NP analysis can be used in order to integrate interactome.

In the last part of the thesis (chapter 4) the proposed pipeline was applied to dynamic transcriptomic datasets collected after exposure to DNA damage which were selected from the literature. The two datasets were using MMS or UV-irradiation as source of the damage.

It was found that both UV-irradiation and MMS induce processes and pathways related to DNA repair, cellular response to heat, vacuolar and proteasomal ubiquitin-dependent protein catabolic processes, response to drug, arginine biosynthesis and autophagy while repressing cell cycle, DNA packaging and ribosome biogenesis.

Doxorubicin was found to be triggering similar processes and pathways with MMS or UV-irradiation. The most striking difference was doxorubicin dependent activation of ribosome biogenesis which might be risen from the time scale of the experiment as discussed in the third chapter.

There are several other differences observed between DNA damage response and transcriptional response to doxorubicin.

It was reported that UV-irradiation triggers nucleotide excision repair (NER) mechanism while oxidative stress or MMS triggers mainly base-pair excision repair (BER) mechanism (Boiteux and Jinks-Robertson, 2013). Interestingly, only mismatch repair (MMR) and NER were identified as responsive process or pathway after exposure to both MMS and UV while doxorubicin was found to be inducing all repair mechanisms; BER, NER and MMR. The genes responsible for this difference were identified to be *OGG1* which specifically excises 7,8-dihydro-8-oxoguanine residues located opposite cytosine or thymine residues in DNA and *CDC9* which is a DNA ligase involve in base excision repair. *OGG1* and *CDC9* were found to be induced in response to doxorubicin whereas they were found to be irresponsive to both MMS and UV-irradiation.

DNA damage was observed to be inducing arginine biosynthesis and this observation is in a good correlation with the reports that tumor cells are arginine dependent (Lind, 2004). Interestingly doxorubicin was identified to be repressing arginine biosynthesis in this study. This could be one of the mechanisms that doxorubicin inhibiting in tumor cells.

Siderophore and iron ion transport processes were found to be induced after 30<sup>th</sup> minute of the doxorubicin addition while these processes were observed to be irresponsive to both MMS and UV-irradiation. It was reported that cardiotoxic effects of doxorubicin might be caused by mitochondrial iron accumulation (Ichikawa *et al.*, 2014).

Tumor cells were also reported to be methionine dependent (Ryu *et al.*, 2011). Supportingly oxidative stress was reported to enhance expression of sulfur assimilation genes (Riboldi *et al.*, 2014) and consequently GSH, cysteine and methionine biosynthesis. The iron-sulfur cluster assembly process was found to be induced in response to DNA damage whereas it was found not to be affected in response to doxorubicin.

Cell wall assembly and ascopore wall assembly processes were also found to be induced after 30<sup>th</sup> minute of the doxorubicin addition while these processes were observed to be irresponsive to both MMS and UV-irradiation.

Autophagy is another affected process by doxorubicin and DNA damage in an opposite fashion. *ATG1* is a serine/threonine kinase which is required for autophagy in yeast and its transcription was observed to be induced ( > 2 fold) upon UV irradiation and drastically

induced ( $> 11$  fold) after exposure to MMS while it was identified to be repressed two-fold after doxorubicin pulse in the present study.

Glucose catabolic process, energy derivation by oxidation of organic compounds and disaccharide biosynthetic process were found to be another set of processes as irresponsive to MMS and UV-irradiation while they were down-regulated upon exposure to doxorubicin.

Interestingly intracellular signaling cascade and cell surface receptor linked signal transduction processes and MAPK signaling pathway were repressed in response to doxorubicin whereas signaling related processes and pathways could not captured as responsive to MMS.

DREM analysis revealed 17 master regulators of response to doxorubicin and 22 transcriptional factors mediating response to DNA damage. Between these two TF sets 10 regulators were identified to be common (Fhl1p, Fkh1p, Hsf1p, Mbp1p, Msn2p, Msn4p, Rpn4p, Skn7p, Swi6p and Yap7p). Abf1p which involves in gene silencing, DNA replication and repair, Ace2p which is a transcription factor required for septum destruction after cytokinesis, Gcn4p which is an activator of amino acid biosynthetic genes, copper-sensing transcription factor Mac1p, pheromone response related Mcm1p, Nrg1p which is a repressor of a variety of processes including filamentous growth and alkaline pH response and Sip4p which is a gluconeogenesis related transcription factor were found to be key regulators of the transcriptomic response to doxorubicin. DNA damage specific TFs were found to be related to iron utilization and homeostasis (Aft2p), stress responses, iron metabolism, and pleiotropic drug resistance (Cad1p and Yap1p), MAPK signaling cascade (Rlm1p, Tec1p, Ste12p and Dig1p), nuclear division (Ndd1p), chromatin silencing and telomere length maintenance (Rap1p) and chromatin remodeling (Ume6p). Some DNA damage specific TFs were observed to be closely related to the common TFs with response to doxorubicin like Fkh2p which is paralog of Fkh1p, Swi4p which is another component of SBF complex with Swi6p.

There was a delay of 30 to 60 minutes in the expected transcriptional response to doxorubicin compared to response to MMS and UV-irradiation. Number of genes whose expression level changes more than two-fold was found to be 402 after treatment with doxorubicin while expression level of 1989 and 216 genes after exposure to the MMS and

UV had changed more than two-fold, respectively. The delay and the difference between magnitudes of the transcriptional responses can be explained by dose which might be unequal to trigger some response mechanisms at a more obvious magnitude. Another reason for these differences might be the different strains which were used in these studies. It should also be noted that experimental design used in these studies are different.

## REFERENCES

Abedin, M. J., D. Wang, M. A. McDonnell, U. Lehmann and A. Kelekar, 2007, “Autophagy delays apoptotic death in breast cancer cells following DNA damage”, *Cell Death & Differentiation*, 14(3), pp. 500-510.

Aguirre, J., D. Papo and J. M. Buldú, 2013, “Successful strategies for competing networks”, *Nature Physics*, 9(4), pp. 230-234.

Albert, R., H. Jeong and A. L. Barabási, 2000, “Error and attack tolerance of complex networks” *Nature*, 406(6794), pp. 378-382.

Anders, C. K., B. Adamo, O. Karginova, A. M. Deal, S. Rawal, D. Darr, A. Schorzman, C. Santos, R. Bash, T. Kafri, L. Carey, C. R. Miller, C. M. Perou, N. Sharpless and W. C. Zamboni, 2013, “Pharmacokinetics and Efficacy of PEGylated Liposomal Doxorubicin in an Intracranial Model of Breast Cancer”, *PLOS One*, 8(5): e61359. doi:10.1371/journal.pone.0061359.

Arga, K. Y., Z. İ. Önsan, B. Kırdar, K. Ö. Ülgen and J. Nielsen, 2007, “Understanding signaling in yeast: insights from network analysis”, *Biotechnology and bioengineering*, 97(5), pp. 1246-1258.

Ashburner M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000, “Gene Ontology: tool for the unification of biology”, *Nature genetics*, 25(1), pp. 25-29.

Assenov, Y., F. Ramírez, S. E. Schelhorn, T. Lengauer and M. Albrecht, 2008, “Computing topological parameters of biological networks”, *Bioinformatics*, 24(2), pp. 282-284.

Baganz, F., A. Hayes, D. Marren, D. C. J. Gardner, and S. G. Oliver, 1997, “Suitability of Replacement Markers for Functional Analysis Studies in *Saccharomyces cerevisiae*”, *Yeast*, 13, pp. 1563-1573.

Bandyopadhyay, S., M. Mehta, D. Kuo, M. K. Sung, R. Chuang, *et al.*, 2010, “Rewiring of genetic networks in response to DNA damage”, *Science*, 330(6009), pp. 1385-1389.

Barabási, A. L. and Z. N. Oltvai, 2004, “Network biology: understanding the cell's functional organization”, *Nature reviews genetics*, 5(2), pp. 101-113.

Baryshnikova, A., M. Costanzo, Y. Kim, H. Ding, J. Koh, K. Toufighi, K., *et al.*, 2010, “Quantitative analysis of fitness and genetic interactions in yeast on a genome scale”, *Nature methods*, 7(12), pp. 1017-1024.

Batada, N. N., L. D. Hurst and M. Tyers, 2006, “Evolutionary and physiological importance of hub proteins”, *PLoS Comput Biol*, 2(7), e88.

Begley, T. J., A. S. Rosenbach, T. Ideker and L. D. Samson, 2004, “Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping”, *Molecular cell*, 16(1), pp. 117-125.

Benton, M. G., S. Somasundaram, J. D. Glasner and S. P. Palecek, 2006, “Analyzing the dose-dependence of the *Saccharomyces cerevisiae* global transcriptional response to methyl methanesulfonate and ionizing radiation”, *BMC genomics*, 7(1), 305.

Benzi M. and C. Klymko, 2015, “On the limiting behavior of parameter-dependent network centrality measures”, *SIAM J. Matrix Anal. Appl.*, 36(2), pp. 686–706.

Bhardwaj, N., P. M. Kim and M. B. Gerstein, 2010, “Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators”, *Science signaling*, 3(146), ra79-ra79.

Binns D., E. Dimmer, R. Huntley, D. Barrell, C. O'Donovan and R. Apweiler, 2009, “QuickGO: a web-based tool for Gene Ontology searching”, *Bioinformatics*, 25:22.

Boiteux, S. And S. Jinks-Robertson, 2013, “DNA repair mechanisms and the bypass of DNA damage in *Saccharomyces cerevisiae*” *Genetics*, 193(4), pp. 1025-1064.

Bolstad, B. M., R. A. Irizarry, M. Astrand, and T. P. Speed, 2003 “A comparison of normalization methods for high density oligonucleotide array data based on bias and variance”, *Bioinformatics*, 19:2, pp. 185-193.

Bonacich, P., 1987, “Power and centrality: A family of measures”, *American journal of sociology*, pp. 1170-1182.

Boyle, E. I., S. Weng, J. Gollub, H. Jin, D. Botstein, J. M. Cherry and G. Sherlock, 2004, “GO:: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes”, *Bioinformatics*, 20(18), pp. 3710-3715.

Brown, J. A., G. Sherlock, C. L. Myers, N. M. Burrows, C. Deng, H. I. Wu *et al.*, 2006, “Global analysis of gene function in yeast by quantitative phenotypic profiling”, *Molecular systems biology*, 2(1).

Burger, K., B. Mühl, T. Harasim, M. Rohrmoser, A. Malamoussi, M. Orban, *et al.*, 2010, “Chemotherapeutic drugs inhibit ribosome biogenesis at various levels”, *Journal of Biological Chemistry*, 285(16), pp. 12416-12425.

Cariaso, M. and G. Lennon, 2012, “SNPedia: a wiki supporting personal genome annotation, interpretation and analysis”, *Nucleic acids research*, 40(D1), D1308-D1312.

Chatr-Aryamontri, A., A. Ceol, L. M. Palazzi, G. Nardelli, M. V. Schneider, L. Castagnoli and G. Cesareni, 2007, “MINT: the Molecular INTERaction database”, *Nucleic acids research*, 35(suppl 1), D572-D574.

Cheung-Ong K., G. Giaever and C. Nislow, 2013, “DNA-damaging agents in cancer chemotherapy: serendipity and chemical biology”, *Chem. Biol.*, 20(5), pp. 648-59.

Ciccia, A. and S. J. Elledge, 2010, “The DNA damage response: making it safe to play with knives” *Molecular cell*, 40(2), pp. 179-204.

Coleman, S. T., E. A. Epping, S. M. Steggerda and W. S. Moye-Rowley, 1999, “Yap1p activates gene transcription in an oxidant-specific fashion”, *Molecular and cellular biology*, 19(12), pp. 8302-8313.

Crucitti, P., V. Latora, M. Marchiori and A. Rapisarda, 2004, “Error and attack tolerance of complex networks” *Physica A: Statistical Mechanics and its Applications*, 340(1), pp. 388-394.

de Hoon M. J. L., S. Imoto, J. Nolan and S. Miyano, 2004, “Open source clustering software”, *Bioinformatics*, 20:9, pp. 1453-1454.

De Keersmaecker S. C. J., I. M. V. Thijs, J. Vanderleyden and K. Marchal, 2006, “Integration of omics data: how well does it work for bacteria?”, *Molecular Biology*, 62(5), pp. 1239-1250.

De Lomana, A. L. G., Q. K. Beg, G. De Fabritiis and J. Villà-Freixa, 2010, “Statistical analysis of global connectivity and activity distributions in cellular networks”, *Journal of Computational Biology*, 17(7), pp. 869-878.

Delaunay, A., D. Pflieger, M. B. Barrault, J. Vinh and M. B. Toledano, 2002, “A thiol peroxidase is an H<sub>2</sub>O<sub>2</sub> receptor and redox-transducer in gene activation”, *Cell*, 111(4), pp. 471-481.

Di Marco A., M. Gaetani and B. Scarpinato, 1969, “Adriamycin (NSC-123,127): a new antibiotic with antitumor activity”, *Cancer Chemother Rep.*, 53(1), pp. 33-37.

Dikicioglu D., E. Karabekmez, B. Rash, P. Pir, B. Kirdar and S. G. Oliver, 2011, “How yeast re-programmes its transcriptional profile in response to different nutrient impulses”, *BMC Systems Biology*, 5:148.

Doherty, K. M., L. D. Pride, J. Lukose, B. E. Snynsman, R. Charles, A. Pramanik, *et al.*, 2012, “Loss of a 20S proteasome activator in *Saccharomyces cerevisiae* downregulates

genes important for genomic integrity, increases DNA damage, and selectively sensitizes cells to agents with diverse mechanisms of action”, *G3: Genes, Genomes, Genetics*, 2(8), pp. 943-959.

Dong, J. and S. Horvath, 2007, “Understanding network concepts in modules”, *BMC systems biology*, 1(1), 24.

Dong, K., S. G. Addinall, D. Lydall J. C. Rutherford, 2013, “The yeast copper response is regulated by DNA damage”, *Molecular and cellular biology*, 33(20), pp. 4041-4050.

Dyer, M. D., T. M. Murali and B. W. Sobral, 2008, “The landscape of human proteins interacting with viruses and other pathogens”, *PLoS Pathog*, 4(2), e32.

Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein, 1998, “Cluster analysis and display of genome-wide expression patterns”, *Proc. Natl Acad. Sci.*, 95, pp. 14863-14868.

Estrada, E., 2006, “Protein bipartivity and essentiality in the yeast protein-protein interaction network”, *Journal of proteome research*, 5(9), pp. 2177-2184.

Estrada, E. and J. A. Rodriguez-Velazquez, 2005, “Subgraph centrality in complex networks” *Physical Review E*, 71(5), 056103.

Ferrarini, L., L. Bertelli, J. Feala, A. D. McCulloch and G. Paternostro, 2005, “A more efficient search strategy for aging genes based on connectivity”, *Bioinformatics*, 21(3), pp. 338-348.

Filkov, V., Z. M. Saul, S. Roy, R. M. D'Souza and P. T. Devanbu, 2009, “Modeling and verifying a broad array of network properties” *EPL (Europhysics Letters)*, 86(2), 28003.

Friedberg, E. C., G. C. Walker, W. Siede, R. D. Wood, R. A. Schultz and T. Ellenberg, 2005, “*DNA Repair and Mutagenesis*”, Second edition, American Society for Microbiology, Washington DC.

Forbes, S. A., G. Bhamra, S. Bamford, E. Dawson, C. Kok, J. Clements, A. Menzies, J. W. Teague, P. A. Futreal and M. R. Stratton, 2008, "The catalogue of somatic mutations in cancer (COSMIC)", *Current protocols in human genetics*, 10-11.

Fornari, F. A., J. K. Randolph, J. C. Yalowich, M. K. Ritke and D. A. Gewirtz, 1994, "Interference by Doxorubicin with DNA Unwinding in MCF-7 Breast Tumor Cells", *Molecular Pharmacology*, vol. 45, no. 4, pp. 649-656.

Freeman, L. C., 1977, "A set of measures of centrality based on betweenness", *Sociometry*, 40:1, pp. 35-41.

Fu, W., B. E. Sanders-Beer, K. S. Katz, D. R. Maglott, K. D. Pruitt and R. G. Ptak, 2009, "Human immunodeficiency virus type 1, human protein interaction database at NCBI", *Nucleic acids research*, 37(suppl 1), D417-D422.

Gasch, A. P., M. Huang, S. Metzner, D. Botstein, S. J. Elledge and P. O. Brown, 2001, "Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p", *Molecular biology of the cell*, 12(10), pp. 2987-3003.

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, *et al.*, 2002, "Functional profiling of the *Saccharomyces cerevisiae* genome", *Nature*, 418(6896), pp. 387-391.

Goh, K. I., M. E. Cusick, D. Valle, B. Childs, M. Vidal and A. L. Barabási, 2007, "The human disease network", *Proceedings of the National Academy of Sciences*, 104(21), pp. 8685-8690.

Greenall, A., G. Lei, D. C. Swan, K. James, L. Wang, H. Peters, *et al.*, 2008, "A genome wide analysis of the response to uncapped telomeres in budding yeast reveals a novel role for the NAD<sup>+</sup> biosynthetic gene BNA2 in chromosome end protection", *Genome Biol*, 9(10), R146.

Hahn, M. W. and A. D. Kern, 2005, "Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks", *Molecular biology and evolution*, 22(4), pp. 803-806.

Hao, D., C. Ren and C. Li, 2012, "Revisiting the variation of clustering coefficient of biological networks suggests new modular structure", *BMC systems biology*, 6(1), 34.

Harper, J. W. And S. J. Elledge, 2007, "The DNA damage response: ten years after", *Molecular cell*, 28(5), pp. 739-745.

Hashemy, S. I., J. S. Ungerstedt, F. Z. Avval and A. Holmgren, 2006, "Motexafin gadolinium, a tumor-selective drug targeting thioredoxin reductase and ribonucleotide reductase", *Journal of biological chemistry*, 281(16), pp. 10691-10697.

Holman, A. G., P. J. Davis, J. M. Foster, C. K. Carlow and S. Kumar, 2009, "Computational prediction of essential genes in an unculturable endosymbiotic bacterium, *Wolbachia* of *Brugia malayi*", *BMC microbiology*, 9(1), 243.

Huang, D. W., B. T. Sherman and R. A. Lempicki, 2009, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources", *Nature Protocols*, 4:1.

Ichikawa, Y., M. Ghanefar, M. Bayeva, R. Wu, A. Khechaduri, S. V. N. Prasad, *et al.*, 2014, "Cardiotoxicity of doxorubicin is mediated through mitochondrial iron accumulation", *The Journal of clinical investigation*, 124(2), pp. 617-630.

Ideker, T. and R. Sharan, 2008, "Protein networks in disease", *Genome research*, 18(4), pp. 644-652.

Jelinsky, S. A., L. D. Samson, 1999, "Global response of *Saccharomyces cerevisiae* to an alkylating agent", *Proceedings of the National Academy of Sciences*, 96(4), pp. 1486-1491.

Jelinsky, S. A., P. Estep, G. M. Church and L. D. Samson, 2000, "Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links

base excision repair with proteasomes”, *Molecular and cellular biology*, 20(21), pp. 8157-8167.

Jensen, L. J., M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, *et al.*, 2009, “STRING 8—a global view on proteins and their functional interactions in 630 organisms”, *Nucleic acids research*, 37(suppl 1), D412-D416.

Jeong, H., S. P. Mason, A. L. Barabási and Z. N. Oltvai, Z. N., 2001, “Lethality and centrality in protein networks”, *Nature*, 411(6833), pp. 41-42.

Jiménez-López, C., J. R. Collette, K. M. Brothers, K. M., Shepardson, R. A. Cramer, R. T. Wheeler and M. C. Lorenz, 2013, “*Candida albicans* induces arginine biosynthetic genes in response to host-derived reactive oxygen species”, *Eukaryotic cell*, 12(1), pp. 91-100.

Jonsson, P. F. and P. A. Bates, 2006, “Global topological features of cancer proteins in the human interactome”, *Bioinformatics*, 22(18), pp. 2291-2297.

Jungmann, J., H. A. Reins, J. Lee, A. Romeo, R. Hassett, D. Kosman and S. Jentsch, 1993, “MAC1, a nuclear regulatory protein related to Cu-dependent transcription factors is involved in Cu/Fe utilization and stress resistance in yeast”, *The EMBO journal*, 12(13), 5051.

Kanehisa, M. and S. Goto, 2000, “KEGG: kyoto encyclopedia of genes and genomes”, *Nucleic acids research*, 28(1), pp. 27-30.

Kang, K. B., C. Zhu, S. K. Yong, Q. Gao and M. C. Wong, 2009, “Enhanced sensitivity of celecoxib in human glioblastoma cells: Induction of DNA damage leading to p53-dependent G1 cell cycle arrest and autophagy”, *Molecular cancer*, 8(1), 1.

Kelley, R. and T. Ideker, 2009, “Genome-wide fitness and expression profiling implicate Mga2 in adaptation to hydrogen peroxide”, *PLoS Genet*, 5(5), e1000488.

Kotamraju, S., C. R. Chitambar, S. V. Kalivendi, J. Joseph and B. Kalyanaraman, 2002, "Transferrin receptor-dependent iron uptake is responsible for doxorubicin-mediated apoptosis in endothelial cells role of oxidant-induced iron signaling in apoptosis", *Journal of Biological Chemistry*, 277(19), pp. 17179-17187.

Langfelder P. and S. Horvath, 2007, "Eigengene networks for studying the relationships between co-expression modules", *BMC Systems Biology*, 1:54.

Langfelder P. and S. Horvath, 2008, "WGCNA: an R package for weighted correlation network analysis", *BMC Bioinformatics*, 9:559.

Leek J. T., E. Mosen, A. R. Dabney and J. D. Storey, 2006, "EDGE: extraction and analysis of differential gene expression", *Bioinformatics*, 22:4.

Li, C. and W. H. Wong, 2001, "Model-based Analysis of Oligonucleotide Arrays: Expression Index Computation and Outlier Detection", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no.1, pp. 31-36.

Lind, D. S., 2004, "Arginine and cancer", *The Journal of nutrition*, 134(10), 2837S-2841S.

MacIsaac, K. D., T. Wang, D. B. Gordon, D. K. Gifford, G. D. Stormo and E. Fraenkel, 2006, "An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*", *BMC Bioinformatics*, 7:113.

McDermott, J. E., D. L. Diamond, C. Corley, A. L. Rasmussen, M. G. Katze and K. Waters, 2012, "Topological analysis of protein co-abundance networks identifies novel host targets important for HCV infection and pathogenesis", *BMC systems biology*, 6(1), 28.

Mewes, H. W., D. Frishman, U. Güldener, G. Mannhaupt, K. Mayer, M. Mokrejs, *et al.*, 2002, "MIPS: a database for genomes and protein sequences", *Nucleic acids research*, 30(1), pp. 31-34.

Milenković, T., V. Memišević, A. Bonato and N. Pržulj, 2011, “Dominating biological networks”, *PloS one*, 6(8), e23016-e23016.

Moye-Rowley, W. S., 2003, “Transcriptional control of multidrug resistance in the yeast *Saccharomyces*”, *Progress in nucleic acid research and molecular biology*, 73, pp. 251-279.

Mustacchi, R., S. Hohmann and J. Nielsen, 2006, “Yeast systems biology to unravel the network of life”, *Yeast*, 23(3), pp. 227-238.

Myers, C., 1998, “The role of iron in doxorubicin-induced cardiomyopathy”, *Seminars in Oncology*, 25(4 Suppl 10), pp. 10-14.

Newman, M. E., 2002, “Assortative mixing in networks”, *Physical review letters*, 89(20), 208701.

Orlowski, R. Z. and D. J. Kuhn, 2008, “Proteasome inhibitors in cancer therapy: lessons from the first decade”, *Clinical Cancer Research*, 14(6), pp. 1649-1657.

Pastor-Satorras, R., A. Vázquez and A. Vespignani, 2001, “Dynamical and correlation properties of the Internet”, *Physical review letters*, 87(25), 258701.

Peri, S., J. D. Navarro, R. Amanchy, T. Z. Kristiansen, C. K. Jonnalagadda, *et al.*, 2003, “Development of human protein reference database as an initial platform for approaching systems biology in humans”, *Genome research*, 13(10), pp. 2363-2371.

Pertea, M. and S. L. Salzberg, 2010, “Between a chicken and a grape: estimating the number of human genes”, *Genome Biol*, 11(5), 206.

Polo, S. E. and S. P. Jackson, 2011, “Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications”, *Genes & development*, 25(5), pp. 409-433.

Pommier, Y. and C. Marchand, 2012, “Interfacial Inhibitors: Targeting Macromolecular Complexes”, *Nature Reviews Drug Discovery*, 11, pp. 25-36.

Raychaudhuri S., J. M. Stuart and R. B. Altman, 2000, "Principal components analysis to summarize microarray experiments: application to sporulation time series", *Pac Symp Biocomput*, 455:466.

Reich M., K. Ohm, M. Angelo, P. Tamayo and J. P. Mesirov, 2004, "GeneCluster 2.0: and advanced toolset for bioarray analysis", *Bioinformatics*, 20:11, pp. 1797-1798.

Schulz, M. H., W. E. Devanny, A. Gitter, S. Zhong, J. Ernst and Z. Bar-Joseph, 2012, "DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data", *BMC Systems Biology*, 6:104.

Riboldi, G. P., C. G. Bierhals, E. P. D. Mattos, A. P. G. Frazzon and J. Frazzon, 2014, "Oxidative stress enhances the expression of sulfur assimilation genes: preliminary insights on the *Enterococcus faecalis* iron-sulfur cluster machinery regulation", *Memórias do Instituto Oswaldo Cruz*, 109(4), pp. 408-413.

Rodriguez-Rocha, H., A. Garcia-Garcia, M. I. Panayiotidis and R. Franco, 2011, "DNA damage and autophagy", *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 711(1), pp. 158-166.

Roy, S., 2012, "Systems biology beyond degree, hubs and scale-free networks: the case for multiple metrics in complex networks", *Systems and synthetic biology*, 6(1-2), pp. 31-34.

Roy, S. and V. Filkov, 2009, "Strong associations between microbe phenotypes and their network architecture", *Physical Review E*, 80(4), 040902.

Ryan, C. J., N. J. Krogan, P. Cunningham and G. Cagney, 2013, "All or nothing: protein complexes flip essentiality between distantly related eukaryotes", *Genome biology and evolution*, 5(6), pp. 1049-1059.

Ryu, C. S., H. C. Kwak, K. S. Lee, K. W. Kang, S. J. Oh, K. H. Lee, *et al.*, 2011, “Sulfur amino acid metabolism in doxorubicin-resistant breast cancer cells”, *Toxicology and applied pharmacology*, 255(1), pp. 94-102.

Sancar, A., L. A. Lindsey-Boltz, K. Ünsal-Kaçmaz and S. Linn, 2004, “Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints”, *Annual review of biochemistry*, 73(1), pp. 39-85.

Seo J. and B. Shneiderman, 2004, “A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections”, *IEEE Symposium on Information Visualization 2004*, pp: 65-72.

Shalem, O., O. Dahan, M. Levo, M. R. Martinez, I. Furman, E. Segal and Y. Pilpel, 2008, “Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation”, *Molecular systems biology*, 4(1), 4.

Song, J. and M. Singh, 2013, “From hub proteins to hub modules: the relationship between essentiality and centrality in the yeast interactome at different scales of organization”, 9(2), e1002910.

Stark, C., B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz and M. Tyers, 2006, “BioGRID: a general repository for interaction datasets”, *Nucleic Acid Research*, 34, D535–D539.

Storey, J. D., W. Xiao, J. T. Leek, R. G. Tompkins and R. W. Davis, 2005, “Significance analysis of time course microarray experiments”, *Proceedings of the National Academy of Sciences of the United States of America*, 102(36), pp. 12837-12842.

Storey J. D., Y. D. James and J. T. Leek, 2007, “The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments” *Biostatistics*, 8(2), pp. 414-432.

Storey J. D., J. T. Leek and A. J. Bas, 2015, “edge: Extraction of differential gene expression”, *R package version 2.2.0*.

Tacutu, R., T. Craig, A. Budovsky, D. Wuttke, G. Lehmann, D. Taranukha, J. Costa, V. E. Fraifeld and J. P. de Magalhaes, 2013, “Human ageing genomic resources: integrated databases and tools for the biology and genetics of ageing”, *Nucleic acids research*, 41(D1), D1027–D1033.

Tamayo P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovski and E. Lander and T. Golub, 1999, “Interpreting patterns of gene expression with self-organizing maps: methods and applications to hematopoietic differentiation”, *Proc. Natl Acad. Sci.*, 96, pp. 2907–2912.

Tansley, S. and K. M. Tolle., 2009, “*The fourth paradigm: data-intensive scientific discovery*”, Vol. 1, Microsoft research , Redmond, WA.

Tsai, R. Y. and T. Pederson, 2014, “Connecting the nucleolus to the cell cycle and human disease”, *The FASEB Journal*, 28(8), pp. 3290-3296.

Tukey, J. W., 1958, “Bias and confidence in not-quite large samples”, *Annals of Mathematical Statistics*, 29:2, pp. 614-614.

Vesanto J. and E. Alhoniemi, 2000, “Clustering of the Self-Organizing Map”, *IEEE Transactions on Neural Networks*, 11:3.

Voorhees, P. M. and R. Z. Orlowski, 2006, “The proteasome and proteasome inhibitors in cancer therapy”, *Annu. Rev. Pharmacol. Toxicol.*, 46, pp. 189-213.

Waddington, C. H., 1942, “Canalization of development and the inheritance of acquired characters”, *Nature*, 150(3811), pp. 563-565.

Wade, S. L., K. Poorey, S. Bekiranov and D. T. Auble, 2009, “The Snf1 kinase and proteasome-associated Rad23 regulate UV-responsive gene expression”, *The EMBO Journal*, 28, pp. 2919-2931.

Wang, H., B. Kakaradov, S. R. Collins, L. Karotki, D. Fiedler, M. Shales, *et al.*, 2009, “A complex-based reconstruction of the *Saccharomyces cerevisiae* interactome”, *Molecular & Cellular Proteomics*, 8(6), pp. 1361-1381.

Wang, J., G. Chen, M. Li and Y. Pan, 2011, “Integration of breast cancer gene signatures based on graph centrality”, *BMC systems biology*, 5(Suppl 3), S10.

Workman, C. T., H. C. Mak, S. McCuine, J. B. Tagne, M. Agarwal, O. Ozier, *et al.*, 2006, “A systems approach to mapping DNA damage response pathways”, *Science*, 312(5776), pp. 1054-1059.

Wuchty, S. and E. Almaas, 2005, “Peeling the yeast protein network”, *Proteomics*, 5(2), pp. 444-449.

Xenarios, I., D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte and D. Eisenberg, 2000, “DIP: the database of interacting proteins”, *Nucleic acids research*, 28(1), pp. 289-291.

Xia K., H. Xue, D. Dong, S. Zhu, J. Wang, Q. Zhang, L. Hou, H. Chen, *et al.*, 2006, “Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms”, *PLoS Computational Biology*, 2:11, e145.

Xia L, L. Jaafar, A. Cashikar and H. Flores-Rozas, 2007, “Identification of genes required for protection from doxorubicin by a genome-wide screen in *Saccharomyces cerevisiae*”, *Cancer Research*, 67(23), pp. 11411- 11418.

Yang, H., J. A. Zonder and Q. P. Dou, 2009, “Clinical development of novel proteasome inhibitors for cancer treatment”, *Expert opinion on investigational drugs*, 18(7), pp. 957-971.

Yıldırım, M. A., K. I. Goh, M. E. Cusick, A. L. Barabási and M. Vidal, 2007, “Drug—target network”, *Nature biotechnology*, 25(10), pp. 1119-1126.

Yu, H., P. M. Kim, E. Sprecher, V. Trifonov M. Gerstein, 2007, “The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics”, *PLoS Comput Biol*, 3(4), e59.

Yuzuak, O., 2012, *Functional analysis of the genes in the glucose metabolic network by a system based modular approach*, M.S., Chemical Engineering Department, Bogazici University.

Zeisel A., A. Amir, W. J. Köstler and E. Domany, 2010, “Intensity dependent estimation of noise in microarrays improves detection of differentially expressed genes”, *BMC Bioinformatics*, 11:400.

Zhang B. and S. Horvath, 2005, “A general framework for weighted gene co-expression network analysis”, *Statistical Applications in Genetics and Molecular Biology*, 4:17.

Zhang, R. and Y. Lin, 2009, “DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes”, *Nucleic acids research*, 37(suppl 1), D455-D458.

Zhu, X., M. Gerstein and M. Snyder, 2007, “Getting connected: analysis and principles of biological networks”, *Genes & development*, 21(9), pp. 1010-1024.

Zotenko, E., J. Mestre, D. P. O’leary and T. M. Przytycka, 2008, “Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality”, *PLoS Comput Biol*, 4(8), e1000140.