Large scale investigation in yeast, to identify novel gene(s)
involved in translation pathway

by

Bahram Samanfar

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Department of Biology,
Ottawa-Carleton Institute of Biology

Carleton University
Ottawa, Ontario

© 2014, Bahram Samanfar

Abstract

As a fundamental step in the gene expression pathway, protein synthesis (also known as translation) plays a crucial role in the biology of a cell. Although much has been discovered about the translation pathway over the last few decades, the list of novel factors affecting the translation pathway continues to grow indicating the presence of other novel players associated with the protein synthesis pathway which are yet to be discovered.

This study aimed to identify novel genes involved in the translation pathway. To this end, we used a variety of large-scale screening techniques followed by low throughput experimental analyses designed for genetic studies in yeast, *Saccharomyces cerevisiae*. Using molecular biology techniques and bioinformatics, we systematically investigated the effect of specific gene deletions on translation fidelity and efficiency, Internal Ribosome Entry Sites (IRESs) functionality and translation-related helicase activities, for a total of ~ 70,000 strain analysis.

We further studied the activity of ~15 genes in more details for their involvement in translation pathway. In light of the current study we propose that there remain other uncharacterized factors that influence translation regulation. Further investigations to characterize these novel factors are recommended.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Dr. Ashkan Golshani for his mentorship, perfect suggestions, encouragement, enthusiastic guidance, supervision and for providing me an inspiring environment to pursue and complete my research successfully. Without him, for sure, some parts of my dreams would not have become a reality.

The members of my thesis committee (Dr. Myron Smith and Dr. Marie-Andree Akimenko) have also been extremely helpful in contributing their wide-ranging experience to my interdisciplinary project. I was extraordinarily fortunate in having Dr. M. Smith on my research committee. Besides his kindness and supports, he advised me on the primacy of using biological questions to motivate my research.

I wish to extend my appreciation to the Department of Biology at Carleton University for their friendship, financial support through this project, and for the access granted to departmental facilities and equipments. I am grateful to Dr. Martin Holcik for the gift of plasmids HSP82-LacZ and HSC82-LacZ as well as his invaluable comments. My closest collaborators and fellow researchers have infused many enthusiastic insights into my research. I have benefited from collaborations with a very powerful bioinformatics team, including Dr. Frank Dehne, Dr. James R. Green, Dr. Sylvain Pitre and Andrew Schonrock for sharing their predicted data, PIPE, and their great help in computational analysis. I also would like to thank Dr. Andrew Emili, University of Toronto, and Dr. Mohan Babu, University of Regina, for providing research materials and for all their detailed and constructive comments.

I am also grateful to many other people who have generously helped me during my PhD research. I want to thank Katayoun Omid, Mohsen Hooshyar, Le Hoa Tan, Kristina Shostak, Daniel Burnside, Imelda Galván Márquez, Megan Sanders, Noor Sunba, Firoozeh Chalabiani, MD Alamgir, Matthew Jessulat, Kama Szereszewski, Anna York-Lyon, Rami Megarbane, Jennifer Yixin Jin, Houman Motesharei, Alex Jesso, Urvi Bhojoo and former colleagues/lab members from the Golshani Lab for their help, support and interest. My special thanks go to Anna York-Lyon, the best undergraduate research assistant in the world, for all her hard working in the lab and proofreading my thesis.

Thanks to Laura Thomas, Lisa Chiarelli, Michelle O'Farrell, Caitlyn McKenzie and Ruth Hill-Lapensee who made all administrative issues very simple. Thanks to Ed Bruggink for his unconditional smile and support. I would like to thank all who were important contributors to the successful completion of this thesis. I apologize in advance to anyone whose name I may have forgotten to mention.

It is difficult to overstate my appreciation and sincere thanks to Dr. Mansoor Omid, my supervisor in Tehran University, who initiated my interest to continue my career in research.

I am especially grateful to my parents, for their care, support and love. Simply, without them I would have not made it, at all.

I wish to thank everyone with whom I have shared various unique experiences in life including Reza Samanfar, Mina Mahzooni and Kianoush Omid; it is a pleasure to convey my gratitude to all of them in my humble acknowledgment.

I would like to thank all my friends especially Alireza, Babak, Maryam, Shaghayegh, Mehdi, Ashkan, Eilnaz, Mohammad, Shabnam, Bardia, Parastoo and Mohsen for all their support and pure friendship.

Last but not the least, I would like to give my special thanks to my beloved wife, Katayoun Omid, whose patience, useful comments, friendship, support and love enabled me to complete this work.

This work was partially funded by an Ontario Graduate Scholarship (OGS).

Statement of contribution

The thesis, “Large scale investigation in yeast to identify novel gene(s) involve in translation pathway” is composed of four studies.

Chapter 2, “A global investigation of gene deletion strains that affect premature stop codon bypass in yeast, *Saccharomyces cerevisiae*”, is the result of experiments primarily designed and carried out by myself. A group of undergraduate and graduate students assisted me and worked under my supervision. Le Hoa Tan, Kristina Shostak and Firoozeh Chalabian helped me with β -galactosidase experiments. Mohsen Hooshyar and Katayoun Omid helped me with SGA and PSA analysis. I wrote the manuscript.

Chapter 3, “The identification and investigation of 7 novel genes involved in IRES-mediated translation in yeast, *Saccharomyces cerevisiae*”, is the result of experiments primarily designed and carried out by myself. A group of undergraduate students including Anna York-Lyon, Kama Szereszewski and Rami Megarbane helped me with β -galactosidase experiments. Kristina Shostak, helped me with SGA and PSA experiments. I wrote the first draft of the manuscript.

Chapter 4, “Efficient prediction of human protein-protein interactions at a global scale”, is a collaborative effort mainly by our computer science collaborator and me. PIPE is designed and performed by Sylvain Pitre and Andrew Schoenrock. PIPE data analysis was performed by me, Sylvain Pitre and Andrew Schoenrock. Biology part of the paper is primarily designed and carried out by myself. Verification of MP-PIPE against experimental data was done by me and Andrew Schoenrock. Translation pathway related experiments were performed mainly by me, with the help of Yuan Gui and Katayoun Omid in β -galactosidase assay. Confirmation of novel molecular markers for seasonal

allergic rhinitis was performed by our collaborator (Mikael Benson) and analyzed primarily by me. Network-wide analysis of hubs and betweenness centrality was carried out by me and Sylvain Pitre. Using the predicted human protein interaction network to assign breast cancer proteins was performed by me and Mohsen Hooshyar. Identification of protein complexes within the human interaction network was done by me and our collaborator, Charles Phillips. I wrote the initial biological related parts of the manuscript.

Chapter 5, ‘Utilizing yeast, genetics to identify novel genes involved in translation-related helicase activity’’, is the result of experiments primarily designed and carried out by me. Anna Lyon-York (undergraduate student) helped me with *β -galactosidase* experiments. I wrote the initial manuscript for this project.

During the course of my PhD studies at Carleton University I have been involved in a variety of interesting projects many of which have been published or submitted in form of a journal publication. The list of my current peer-reviewed publications is presented in Appendix B.

Table of Contents

Abstract	ii
Acknowledgements	iii
Statement of contribution	v
Table of contents	vii
List of tables	xii
List of figures	xiii
List of appendices	xv
List of abbreviations	xvii
1 Chapter: Introduction	1
1.1 Systems biology	1
1.1.1 Functional genomics	5
1.1.2 Yeast functional genomics	5
1.2 Protein-protein and genetic interactions	8
1.3 Translation pathway	15
1.3.1 Ribosome	18
1.3.2 Translation activation	19
1.3.3 Translation initiation	21
1.3.4 Translation elongation	22
1.3.5 Translation termination	23
1.4 Internal Ribosome Entry Site (IRES)	24
1.5 Translation process and diseases	27
1.6 Objective	29

2 Chapter: A global investigation of gene deletion strains that affect premature stop codon bypass in yeast, <i>Saccharomyces cerevisiae</i>	31
2.1 Abstract	31
2.2 Introduction	31
2.3 Materials and methods	33
2.3.1 Strains	33
2.3.2 Media and antibiotics	33
2.3.3 Plasmids	33
2.3.4 Gene knockout	34
2.3.5 Primers	34
2.3.6 qRT-PCR	34
2.3.7 Genetic materials extractions	34
2.3.8 <i>β</i> -galactosidase assay	35
2.3.9 Spot test (drug sensitivity analysis)	35
2.3.10 SGA and PSA analysis	35
2.4 Results and discussion	36
2.4.1 Identification of gene deletions that affect expression of <i>β</i> -galactosidase gene with premature stop codons	36
2.4.2 <i>OLA1</i> , <i>BSC2</i> and <i>YNL040W</i> affect protein biosynthesis	41
3 Chapter: The identification and investigation of 7 novel genes involved in IRES-mediated translation in yeast, <i>Saccharomyces cerevisiae</i>	54
3.1 Abstract	54
3.2 Introduction	54

3.3	Material and methods	57
3.3.1	Yeast strains, media and plasmids	57
3.3.2	Gene knockout and DNA transformation	57
3.3.3	Primers	57
3.3.4	Genetic material extractions	57
3.3.5	qRT-PCR	58
3.3.6	β-galactosidase assay	58
3.3.7	Spot test (Drug sensitivity analysis)	58
3.3.8	SGA and PSA analysis	58
3.4	Results and Discussion	59
3.4.1	Identification of novel candidate genes affect IRES-mediated translation pat	59
3.4.2	Genetic interaction (SGA and PSA) investigations	66
4	Chapter: Efficient prediction of human protein-protein interactions	
	at a global scale	78
4.1	Abstract	48
4.2	Introduction	48
4.3	Materials and methods	83
4.3.1	Biological experiments	83
4.3.1.1	Yeast growth condition	83
4.3.1.2	Drug sensitivity test	83
4.3.1.3	Protein expression assay	83
4.3.1.4	qRT-PCR	84
4.3.1.5	Versatile affinity-tagging, purification, and protein identification	85

4.3.1.6	Prediction and analysis of candidate biomarkers for seasonal allergy	86
4.3.2	Computational	87
4.3.2.1	Sequential PIPE algorithm	87
4.3.2.2	MP-PIPE overview	89
4.3.2.3	Complete scan of the human proteome	92
4.3.2.4	Hubs and betweenness centrality	93
4.3.2.5	Graph algorithms for assigning protein complexes	94
4.4	Results and discussion	96
4.4.1	MP-PIPE performance and scalability enables computational scan of entire human proteome	96
4.4.2	Verification of MP-PIPE against experimental data	97
4.4.3	All-against-all (pair-wise) scan of the human proteome	98
4.4.4	The proteome-wide PPI network can identify translation genes	111
4.4.5	Identification of novel molecular markers for seasonal allergic rhinitis	115
4.4.6	Network-wide analysis of hubs and betweenness centrality	118
4.4.7	Using the predicted human protein interaction network to assign breast cancer proteins	122
4.4.8	Identification of protein complexes within the human interaction network	126
5	Chapter: Utilizing yeast, genetics to identify novel genes involved in translation-related helicase activity	131
5.1	Abstract	131
5.2	Introduction	131
5.3	Materials and methods	133

5.3.1	Media and strains	133
5.3.2	Plasmids	134
5.3.3	DNA transformation	134
5.3.4	<i>β-galactosidase</i> assay	134
5.4	Results and discussion	135
6	Chapter: Conclusion	143
6.1	Conclusion	143
6.2	Closing remarks and future directions	152
	References	154
	Appendices A	178
	Appendices B	214

List of tables

Table 1-1	Possible wobble-pairs at the 1st, 2nd and 3rd bases at mRNA→ tRNA level_____	<u>20</u>
Table 1-2	Gene Ontology analysis on IRES containing genes_____	<u>26</u>
Table 4-1	Percentages of <i>H. sapiens</i> pairs in which both partners share the same GO, gene ontology, annotation as well as third party interactions_____	<u>101</u>
Table 4-2	Overlap of co-purifying proteins identified through LGTS-MS with previously reported (known) interactions and MP-PIPE predictions ____	<u>109</u>
Table 4-3	Enrichment of biological process for proteins with highest centrality measurements; hubs and betweenness centrality (top 50)_____	<u>120</u>
Table 5-1	Normalized β -galactosidase assay analysis_____	<u>139</u>

List of figures

Figure 1-1	An overview of systems molecular biology_____	4
Figure 1-2	Concept of overlapping pathways_____	13
Figure 1-3	Eukaryotic translation process_____	17
Figure 2-1	A representative yeast lift assay_____	38
Figure 2-2	Drug sensitivity analysis_____	43
Figure 2-3	A) <i>β-galactosidase</i> expression analysis. B) <i>β-galactosidase</i> mRNA content analysis_____	44
Figure 2-4	A) Translation rate measurement. B) <i>β-galactosidase</i> mRNA content analysis_____	45
Figure 2-5	SGA analysis_____	49
Figure 2-6	Representative spot test confirmation for phenotypic suppression analysis of <i>OLAI</i> _____	51
Figure 3-1	Spot test analysis for selected candidates under acetic acid condition_____	61
Figure 3-2	<i>HSP82</i> and <i>HSC82</i> mRNA content analysis_____	62
Figure 3-3	A) <i>β-galactosidase</i> expression from templates that contain an IRES element fused to the LacZ expression cassette. B) <i>β-galactosidase</i> mRNA content analysis_____	65
Figure 3-4	The representative SGA analysis for <i>VPS70</i> _____	69
Figure 3-5	Re-introduction of deleted target genes in double mutants_____	71
Figure 3-6	Representative spot test confirmation for phenotypic suppression analysis of <i>VPS70</i> _____	73

Figure 4-1	Distribution of the interacting protein pairs on the basis of subcellular localization (A), molecular function (B) and cellular process (C) for both previously detected interactions and interactions unique to this study normalized by the number of possible pairs with both GO terms_____	102
Figure 4-2	Affinity purification experiments using P83916 (CBX1), Q99496 (RNF2), P16104 (H2AFX), and Q09028 (RBBP4) as baits_____	105
Figure 4-3	A) Number and B) percentage of co-localized interacting pairs by GO components in previously reported data compared with those reported in this study only _____	110
Figure 4-4	The relative <i>β-galactosidase</i> activity. B) The relative mRNA level C) Increased sensitivity of <i>hap2Δ</i> , <i>erg28Δ</i> and <i>lys12Δ</i> to different translation inhibitory drugs D) Effect of gene deletions on translation efficiency_____	113
Figure 4-5	Analysis of candidate proteins with ELISA_____	117
Figure 4-6	Comparing the number of disease-associated proteins with high degree (hubs) and high betweenness centrality_____	121
Figure 4-7	Schematic diagram of breast cancer pathway_____	124
Figure 4-8	Schematic representation of the interactions for three putative complexes_____	129
Figure 5-1	A representative yeast gene deletion colony array subjected to X-gal lift assay_____	137
Figure 5-2	A) Gene Ontology annotation analysis. B) Interaction map_____	140

List of appendices

Appendices-A

App. 1-1	list of available yeast resources (database) derived from large-scale studies_____	178
App. 2-1	List of designed primers for gene knock-out in yeast_____	179
App. 2-2	Alphabetically ordered list of genes identified through large-scale X-gal based <i>β-galactosidase</i> assay for stop codon read-through_____	180
App. 2-3	SGA results of synthetic sick interactions for Ola1, Bsc2 and YNL040W_____	183
App. 2-4	Re-introductions of deleted target genes in double mutants_____	184
App. 3-1	list of primers_____	186
App. 3-2	SGA analysis_____	187
App. 3-3	Re-introduction of deleted target genes in double mutants_____	194
App. 3-4	List of gene deletion mutants compensated by overexpression plasmids in different stress condition_____	195
App. 4-1	Precision-Recall curve for <i>H. sapiens</i> PPIs_____	199
App. 4-2	MP-PIPE benchmark performance_____	200
App. 4-3	List of <i>Homo sapiens</i> interactions predicted in this study_____	201
App. 4-4	Number interacting pairs where both proteins have the same function_____	204
App. 4-5	List of interactions for fibroblast growth factors (FGFs), cyclin- dependent kinases (CDKs), FGF regulators and CDK inhibitor/activators_____	205

App. 4-6	List of proteins from the acute phase response pathway, complement pathway and glucocorticoids receptor pathway_____	<u>206</u>
App. 4-7	List of hub proteins in <i>Homo sapiens</i> ranked by the number of neighbors for each protein_____	<u>207</u>
App. 4-8	List of betweenness centrality proteins in <i>Homo sapiens</i> ranked by highest betweenness centrality measure_____	<u>212</u>
App. 4-9	List of protein mutations associated with resistance to breast cancer__	<u>213</u>
Appendices-B	List of publications_____	<u>214</u>

List of abbreviations

μg	Micro grams
μL	Micro liters
°C	Degrees Celsius
3-AT	3-Amino-1,2,4 triazole
AD	Activation domain
APMS	Affinity purification combined with mass spectrometry
ATP	Adenosine triphosphate
CBP	Calmodulin binding protein
cDNA	Complementary DNA
clonNAT	Nourseothricin
DB	DNA binding domain
DNA	Deoxyribonucleic acid
ECL	Enhanced chemiluminescence
eEF	Eukaryotic elongation factor
eIF	Eukaryotic initiation factor
EGTA	Ethylene glycol tetra-acetic acid
g	Gram
G418	Geneticin

GAL	Galactose
GD	Growth detector
GI	Genetic interaction
IRES	Internal ribosome entry site
ITAFs	IRES transacting factors
LB	Lysogeny broth
LGTS-MS	Lentivirus-delivered gateway-compatible affinity tagging system coupled with mass spectrometry
MAT	Mating type locus
mg	Milligram
min	Minute
mL	Milli liter
mM	Milli molar
MP-PIPE	Massively parallel protein-protein interaction prediction engine
mRNA	Messenger RNA
MS	Mass spectrometry
ng	Nanogram
OD	Optical density
ONPG	O-nitro-phenyl- β -D-galactoside
ORF	Open reading frame

PABP	Poly A binding protein
PCD	Programmed cell death
PCR	Polymerase chain reaction
PIPE	Protein-protein interaction prediction engine
PKA	Protein kinase A pathway
PPI	Protein-protein interaction
PSA	Phenotypic suppression analysis
qRT-PCR	Quantitative RT-PCR
rDNA	Ribosomal DNA
RF	Release factor
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT	Room temperature
RT-PCR	Reverse transcriptase PCR
SAR	Seasonal allergic rhinitis
SC	Synthetic complete media
SDL	Synthetic dosage lethality
SDS-PAGE	Sodium dodecyl sulfate poly-acrylamide gel electrophoresis
SGA	Synthetic genetic array

SGD	Saccharomyces genome database
SUI	Suppression of initiation codon
TAE	Translation associated element
TAP	Tandem affinity purification
TC	Ternary Complex
TEV	Tobacco etch virus
TOR	Target of rapamycin
tRNA	Transfer RNA
UAS	Upstream activation sequence
URA	Uracil
UTR	Untranslated region
UV	Ultraviolet
WT	Wild type
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
Y2H	Yeast two hybrid
YPD	Yeast extracts peptone dextrose media

1 Chapter: Introduction

1.1 Systems biology

For more than a century, scientists have been investigating the biology of genes and genetic information using different model organisms. Such organisms, including *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, etc., have been used to gain a better understanding of inheritable mechanisms and cellular processes that govern cell responses to internal and external stimuli (Castrillo and Oliver, 2004). Generally speaking, classical genetic studies investigate a single gene or a gene family, one at a time, and try to combine isolated data points to have a better understanding the biology of a cell. These approaches are resource intensive and time-consuming and, in most cases, do not accurately represent the biology of the whole system due to a lack of comprehensive coverage and the absence of clear data for global communications between cellular compounds and pathways. These limitations have inspired the development of new techniques that give scientists the ability to study a number of genes, proteins and pathways at a same time. Collectively, these methods are termed large-scale investigations and high-throughput methods. A growing number of high-throughput techniques are being developed to address various challenges and questions in the field of genetics (Rigaut *et al.*, 1999; Tong *et al.*, 2001; Parson *et al.*, 2004; Hillenmeyer *et al.*, 2008).

For the last two decades, an increase in the availability of the genome-wide DNA sequences of various model and non-model organisms has significantly impacted the field of molecular biology and genetics. During this period, a variety of large-scale

analyses combined with improved high-throughput methods have been developed and applied in order to study the biology of genes and genetics. These methods and strategies investigate not only the basic function of genes, but also their cross-communications with different pathways at transcriptome, proteome, and metabolome levels (Oliver *et al.*, 2002).

In the context of large-scale studies, the budding yeast, *S. cerevisiae*, has been heavily utilized for investigating the biology of eukaryotic cells at a systems level. *S. cerevisiae* has been subjected to various large-scale experiments such as genome sequencing (Goffeau *et al.*, 1996), expression profiling (Ghaemmaghami *et al.*, 2003) and interaction mapping (Krogan *et al.*, 2006; Alamgir *et al.*, 2008; Costanzo *et al.*, 2010; Samanfar *et al.*, 2013; 2014; Omid *et al.*, 2014). A representative list of available yeast resources derived from such large-scale investigations is presented in appendix 1-1.

The systematic investigation of cellular processes at the molecular level is often referred to as systems molecular biology. It presents an integrated approach using genetics, molecular genetics and genomics and bioinformatics to explain functional relationships between genes and other cellular components from a sub-cellular to an organism level as a whole system. For the most part, systems molecular biology is performed by an interdisciplinary team of molecular and computational biologists. It focuses not only on the importance of global and integrative modern biological investigations, but also on the development of novel and improved approaches to study the complexity of a cell's biology (Aderem, 2005; Sopko *et al.*, 2006; Pitre *et al.*, 2008a; 2008b). The main goal of systems molecular biology is to develop testable hypotheses and to have comprehensive understandings of life at a molecular networks level (Castrillo and Oliver, 2004) (Figure

1-1). Some computational investigations in systems molecular biology aim to predict interaction networks within a system; others focus on integrating the available data obtained from different high-throughput techniques to computationally simulate, analyze or predict cellular responses (Jessulat *et al.*, 2011).

Figure 1-1: An overview of systems molecular biology. Systems molecular biology mediates the development of hypotheses using various approaches to investigate and integrate pathways, networks and interactions to have a better understanding of how a cell functions.

1.1.1 Functional genomics

Around 1997, the term "Functional Genomics" emerged to describe investigations that focus on understanding genes and their functions in a systematic manner (Hieter and Boguski, 1997). Nowadays, it also covers comprehensive genome-wide approaches and analytical tools to identify and characterize novel genes and to study gene expression regulations at molecular levels (Parsons *et al.*, 2006). In recent years, the output of a number of high-throughput techniques such as RNAi-based approaches (Tischler *et al.*, 2008), DNA microarrays (Moffat and Sabatini, 2006), protein arrays (Ruffner *et al.*, 2007), Synthetic Genetic Array analysis (SGA) (Tong *et al.*, 2001) and so on, have made a significant contribution to the advancement of this field.

One of the most challenging aspects of systems biology investigations is the analysis and management of the complex data derived from large-scale studies. This makes their interpretation and analysis a difficult task for molecular biologists. As such, certain computational approaches have been developed to help visualize, understand and interpret such complicated data sets. A drawback of these computational tools is that they often suffer from high rates of false positive/negative data points (Delneri *et al.*, 2001; Oliver *et al.*, 2002; Phelps *et al.*, 2002; Urbanczyk-Wochniak *et al.*, 2003; Venancio *et al.*, 2010; Jessulat *et al.*, 2011).

1.1.2 Yeast functional genomics

The yeast, *S. cerevisiae*, is the first eukaryotic model organism used for large-scale genetic analysis (Suter *et al.*, 2006). Genome-wide analysis such as expression array (DeRisi *et al.*, 1997), mass spectrometry analysis for protein-protein interaction (Ho *et al.*, 2002), protein array (Zhu *et al.*, 2001), synthetic lethality screens (Tong *et al.*, 2001)

and chemical genetic analysis (Alamgir *et al.*, 2008) were first validated in yeast using well-controlled reproducible conditions. The large amount of information gathered for more than a decade from fully sequenced genome of yeast, in combination with the availability of various high-throughput resources including the collection of non-essential and essential gene deletion arrays (Winzeler *et al.*, 1999) and yeast gene overexpressed array (Zhu *et al.*, 2001; Sopko *et al.*, 2006), gives *S. cerevisiae* a unique advantage as a model organism for systems molecular biology investigations. The development of the yeast non-essential gene deletion array (Winzeler *et al.*, 1999; Giaever *et al.*, 2002; Suter *et al.*, 2006) is one of the major breakthroughs in biological sciences. In this array, each of the open reading frames (ORFs) is systematically deleted and replaced with *kanMX4* (using homologous recombination ability of the yeast), a kanamycin resistance cassette. Drug sensitivity profiling, Synthetic Genetic Array (SGA), Synthetic Dosage Lethality (SDL) and Phenotypic Suppression Array (PSA) analysis can now be readily performed in yeast in a cost-effective manner (Tong *et al.*, 2001; Sopko *et al.*, 2006; Alamgir *et al.*, 2008; Omid *et al.*, 2014; Samanfar *et al.*, 2013; 2014).

An interesting application of gene deletion array (single or double deletion) is chemical genetic analysis (Galván Márquez *et al.*, 2008; 2013). Chemical genetics is often used in drug discovery but it can also be used in molecular mechanism determination studies (Alamgir *et al.*, 2008; 2010). Due to recent advances in the applicability of chemical genetics in yeast, chemical genome profiles of hundreds of chemicals are now available (Hoepfner *et al.*, 2014; Lee *et al.*, 2014). Altogether, data gathered through these techniques have helped explain the regulatory mechanism of genes, proteins, metabolites, networks and pathways in yeast. Utilizing such a vast amount of data in yeast has offered

fundamental insights into the complexity of biological networks in this microorganism (Lesage *et al.*, 2004; Omid *et al.*, 2014; Samanfar *et al.*, 2013; 2014). As of today, many novel and fundamental genetic interactions have been identified using the *S. cerevisiae* gene deletion arrays. These interactions have provided important functional information about the complexity of pathway communications in higher eukaryotic organisms such as human (Lehner, 2007; Samanfar *et al.*, 2013; 2014). Similarly, the yeast overexpression array has also opened another era of functional genomics to investigate functional pathways and processes within a cell (Boone, 2007). It is now possible to systematically evaluate cell responses to various stimuli, not only in the absence of a gene product, but also when that gene product is overexpressed.

Generally speaking, evolutionary analysis can provide valuable information about conserved and non-conserved sequences of the genome and their structure and function (Glaser and Boone, 2004) that may, in turn, lead to a better understanding of various cellular processes in different organisms. With human sequencing data available, it has been observed that there is ~30% homology between yeast and human genes (Suter *et al.*, 2006) and that nearly ~50% of genes related to human genetic disorders have yeast orthologs (Hartwell, 2004). The presence of such common features between human and yeast further highlights the importance of studying yeast as a model organism to investigate human diseases. In this context, yeast orthologs have been identified for a number of human genes involved in different genetic disorders including neurodegenerative diseases such as Parkinson's disease (PKD) (Outeiro and Giorgini, 2006), Alzheimer disease (Bagriantsev and Liebman, 2006) and Aging (Petranovic and Neilsen, 2008). Although the molecular mechanisms between yeast and more complex

eukaryotes can have notable differences, experimental evidences suggest that many fundamental processes are conserved between these organisms. This has led to the use of yeast genetics as a fundamental platform to study human genetics (Boube *et al.*, 2002; Samanfar *et al.*, 2013). Despite the availability of a variety of large-scale methodologies developed for yeast genetics, there still remains a demand for improved and more reliable methodologies as well as new computational tools that can help produce data with improved efficacy and improved data result interpretation.

1.2 Protein-protein and genetic interactions

The term “interaction” in the context of two gene products can be divided into two main categories of Protein-Protein Interaction (PPI) and Genetic Interaction (GI). PPI happens when two or more proteins physically interact with each other to perform a specific function (Butland *et al.*, 2007). It is now generally accepted that proteins perform their functions by physically interacting with each other, forming transient or stable complexes. It is these complexes that are explained as functional units within a cell (Pitre *et al.*, 2006; Jessulat *et al.*, 2011). This contrasts with GIs that are explained with the concept of overlapping pathways that compensate for each other’s activities. In the absence of a functional pathway (e.g. *via* mutation in a corresponding gene), a parallel pathway can kick in and compensate for the missing function with no or little phenotypic consequence. However, when both pathways are compromised a new phenotype is observed (Tong *et al.*, 2001; Jessulat *et al.*, 2011; Omid *et al.*, 2014).

In biological systems, investigating the details of biochemical pathways and elucidating the interactions among them is an important and challenging task. With the advent of improved large-scale techniques, it is presently possible (to an extent) to systematically

analyze and investigate the fundamentals of certain pathways using protein-protein and genetic interaction analysis in a high-throughput fashion. Investigating PPIs and GIs is a powerful approach to study the relation between genes/proteins, various pathways and processes. It helps unravel some of the network interactions and high level communications between pathways that govern the biology of a cell.

Yeast Two-Hybrid (Y2H) (Fields and Song, 1989) and Tandem Affinity Purification (TAP) coupled with Mass Spectrometry (MS) (Rigaut *et al.*, 1999) are two established approaches to identify PPIs in a genome-wide manner. Application of these methodologies has generated thousands of functional associations between proteins in different cell lines (Stelzl and Wanker, 2006).

Briefly, Y2H relies on reconstitution of transcriptional activator proteins. These proteins contain two functional domains; a DNA-binding Domain (DB) that binds at the Upstream Activation Sequence (UAS) and a transcriptional Activator Domain (AD) that is responsible for transcription machinery recruitment. In a given Y2H experiment, BD and AD regions are separately fused to two proteins of interest, X and Y, to form BD-X and AD-Y hybrid proteins. If X and Y have an affinity for each other and physically interact, then BD and AD regions will be brought together allowing them to become functional once again and activating the expression of a reporter gene. Initially, Field and Song (1989) used the BD and AD of GAL4 transcription factor and the LacZ expression cassette that encodes for *β -galactosidase*. After the formation of UAS-BD-AD physical complex, the LacZ reporter system was activated and produced *β -galactosidase* enzyme that was quantifiable using standard X-gal-based *β -galactosidase* assay. Other Y2H

cassettes are now available that are based on the ability of yeast to grow under different growth conditions (Caufield *et al.*, 2012; Rajagopala *et al.*, 2012).

TAP-MS approach is based on a double affinity tagging of a target protein followed by MS identification of co-purified proteins. The TAP-tag consists of Calmodulin Binding Protein (CBP), and a Tobacco Etch Virus (TEV) cleavage site and *Staphylococcus* protein A that binds to the IgG domain. After cell lysis, the tagged protein and its corresponding complex are purified through two-step affinity purification. In the first step, the tagged protein binds to the IgG coated beads; the protein is released from the beads using TEV protein digestion and then subjected to a second purification step using calmodulin beads. Proteins are then released from these beads using Ethylene Glycol Tetra-acetic Acid (EGTA). This eluent, consisting of the target protein and its interacting partners, is then subjected to Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) to separate different protein partners. Each of the interacting partners is then identified through MS analysis (Rigaut *et al.*, 1999).

Although experimental techniques such as Y2H and TAP have provided a vast quantity of information regarding PPIs, they remain expensive and labor intensive (Skrabanek *et al.*, 2008). In addition, they have inherent technical disadvantages making them very difficult to use to study certain proteins; these limitations are highlighted by an absence of a good overlap between the data obtained by different techniques (Gomez *et al.*, 2003; Szilagyi *et al.*, 2005; Jessulat *et al.*, 2011). Having said that, it is imperative to develop reliable computational methods to facilitate accurate predictions and identifications of PPIs to bridge the gap of comprehensive knowledge regarding interaction networks (Guo *et al.*, 2008).

Over the past decade, a number of computational methods capable of predicting a subset of cellular PPIs have been developed. Some of these methods are based on previously identified domains (Kim *et al.*, 2002; Han *et al.*, 2004), some are designed to evaluate the evolutionary relationship between interacting proteins (Huang *et al.*, 2004; Espadaler *et al.*, 2005) while others use the structural information of proteins to predict physical interactions (Aloy and Russell, 2003; Ogmen *et al.*, 2005). It has been previously shown that short re-occurring polypeptide sequences can be used to predict novel PPIs (Martin *et al.*, 2005; Pitre *et al.*, 2006). Computational prediction methods based on this observation have the advantage of being able to make predictions using only the primary sequence of proteins instead of other features such as 3D structure information. In this way, they are independent of other features that may not be available to a notable portion of the proteomes. The Protein-protein Prediction Interaction Engine (PIPE) (Pitre *et al.*, 2006; 2008a; 2008b; 2012) makes use of these short re-occurring polypeptide sequences found in experimentally detected interactions to make new PPI predictions. It was originally designed and tested using the model organism *S. cerevisiae* (2006, 2008) and it has since been extended to other organisms including *Caenorhabditis elegans* and *Schizosaccharomyces pombe* (2011). Since PIPE relies on primary structure only, it can be used on proteins that do not have additional annotations required by most other computational PPI prediction methods. Beside traditional *in vivo* methods such as Y2H and TAP-tagging where some proteins cannot be tested due to technical limitations, PIPE can make predictions on any proteins (with a known sequence) since this is done *in silico*.

GIs take place when two or more genes interact through biologically overlapping pathways (Figure 1-2). These overlapping pathways can compensate for the absence of

each other. Generally, when the phenotypic consequence of a double gene deletion cannot be explained by the individual phenotypes of single gene deletions, it is said that the two genes are genetically interacting. Similarly, if the individual phenotypic consequences for the deletion of one gene and overexpression of another cannot explain the phenotypic consequence of combined deletion of one and overexpression of the other, these two genes are again thought to be genetically interacting. GI often confers a functional relationship (Tong *et al.*, 2001; 2004) and can explain some of the complex association of networks that govern intracellular cross-talks between different cellular processes (Krogan *et al.*, 2006; Costanzo *et al.*, 2010; Omid *et al.*, 2014).

Synthetic Genetic Array (SGA) analysis is a method developed to screen the fitness of double gene deletion yeast mutants in a systematic manner. It is a large-scale method used to generate double deletions and to score their fitness using colony size growth.

Figure 1-2: Pathways A & B are assumed to be overlapping pathways. Mutation of a gene in one pathway can be compensated by a parallel (overlapping) pathway with no phenotypic effect. Mutation in two genes in both overlapping pathways will produce a synthetic lethality or sickness phenotype that is not readily explained by the phenotype of individual gene deletions. In this example, genes associated with pathway A often form genetic interactions with those of pathway B.

SGA method in yeast is based on a theory that considers ~80% of yeast genes have at least one genetic partner with an overlapping function (Tong *et al.*, 2001; Hillenmeyer *et al.*, 2008). When a partner gene is deleted, no specific phenotype will be observed due to compensation activity by the other partner gene. However, inactivation of both genetic partners (double mutation) will often render a phenotypical consequence that is detectable by lethality (absence of colony for double mutant colonies) or sickness (decrease in colony size for double mutant colonies); these indicate negative GI (Figure 1-2) (Tong *et al.*, 2004). In contrast, for positive GI double deletion colonies show an increased level of phenotypic fitness in comparison to corresponding single deletions.

In SGA, a query strain carrying a mutation for the gene of interest (deleted and replaced with a selectable marker gene in mating type α) is systematically mated with a non-essential haploid deletion set (~5000 mutants in mating type a). After a few rounds of selections, a progeny of double mutants is obtained, with each strain carrying the query mutation in association with a second deletion (double mutants). The measured fitness of double mutants is then used to identify GIs (Tong *et al.*, 2001).

Synthetic Dosage Lethality (SDL) (Sopko *et al.*, 2006) and Phenotypic Suppression Analysis (PSA) (Alamgir *et al.*, 2008) are two modified versions of SGA technique. SDL is used to investigate gene dosage effect (overexpression of one gene and deletion of other gene). PSA uses the same concept as SDL but investigates the overexpression of one gene compensating the absence of a potential partner gene in overlapping pathways under a specific stress condition.

1.3 Translation pathway

Genes are inheritable materials, stored in the form of DNA, which are transcribed and generally translated into proteins. In eukaryotes, a gene is transcribed from the DNA template strand by RNA polymerase to produce a complementary RNA strand called the pre-mRNA inside the nucleus. The pre-mRNA is modified into a mature mRNA by the removal of non-coding regions (introns), the addition of a cap structure at the 5' end and the attachment of a poly(A) tail at the 3' end. These modifications are necessary for the recruitment of ribosomes, the protection of the mRNA from degradation, and for efficient translation regulation (Gallie, 1991). Modified mRNA is then transported from the nucleus to the cytoplasm where translation starts. Translation is the step at which genetic information becomes meaningful and functional products are produced. Due to its central importance, many antibiotics target this process. In brief, translation refers to the biological processes that leads to polymerization of amino acids into polypeptide chains from the genetic codes (DNA) embedded in mRNA molecules. mRNA molecules contain triplet genetic codons that call for specific amino acids. Ribosomes are one of the most important macromolecules in a cell. They are responsible for synthesizing new proteins by reading genetic codes carried by mRNA molecules.

In eukaryotes, the smaller subunit of the ribosome binds to 5' end of the mRNA (Cap). The two subunits unite near the initiation codon to form a complete ribosome capable of mediating the synthesis of polypeptide chains. When the translation process is completed, the complex of ribosomal large and small subunits and mRNA will disassemble. Translation of genetic code additionally requires the help of an adaptor molecule called tRNA. tRNA contains three nucleotides complementary to the codon, called the anti-

codon. Another region of tRNA is covalently bonded to its corresponding amino acid. Inside the ribosome, hydrogen binding of the tRNA and mRNA holds the amino acids in a proper location within a ribosome where peptide bonds are formed. This process continues as the ribosome moves through the mRNA and forms polypeptide chains (Kozak, 2001; 2002; 2005; 2007). Classically, translation has four (one + three) major steps: activation, initiation, elongation and termination (Figure 1-3).

Figure 1-3: Eukaryotic translation process. Translation initiation is the step at which ribosome subunits, in association with translation initiation factors, initiate translation at the AUG (start) codon position. The ribosome (in association with translation elongation factors and tRNAs) will then move towards the 3' end of the mRNA, reading genetic codes. In this step, the polypeptide chain is synthesized. Finally, at the termination step (stop codon detection), release factors disassemble ribosome to its subunits and the polypeptide chain (immature protein) is released.

1.3.1 Ribosome

Eukaryotic ribosomes are Mega-Dalton molecules of around 25-30 nm in diameter with rRNA/protein ratio of around 1. Ribosome biogenesis is a highly complex and conserved process that includes rRNA transcription, modification, processing, folding and assembly. Ribosomal RNA (rRNA) transcription accounts for ~ 55-60% of total transcription in a cell; up- and down-regulation of ribosome biogenesis can have a substantial effect on the process of gene expression (Rudra and Warner, 2004). The majority of our current knowledge regarding ribosome biogenesis comes from studies on yeast, where a large number of ribosome precursors and components have been identified and characterized (Granneman and Baserga, 2004). Ribosomes have two subunits; one large and one small. In bacteria, they are 30S (small) and 50S (large) free subunits; the 70S ribosome refers to a combined stage of ribosome (whole ribosome). In eukaryotes such as yeast, ribosomes consist of 40S and 60S free subunits and the 80S is the whole ribosome. The eukaryotic small subunit contains 18S rRNA and 30-50 subunit proteins (Planta and Mager, 1998). The large subunit contains 25S, 5.8S and 5S rRNAs and 40-50 subunit proteins (Fromont-Racin *et al.*, 2003; Nazar, 2004).

In eukaryotes, production of rRNA is loosely associated with the nucleolus. There, RNA polymerase I transcribes rDNA as a polycistronic transcript known as pre-rRNAs. These pre-rRNAs contain external and internal transcribed spacers along with the mature rRNA (5.8S, 18S, 25-38S rRNA). The pre-rRNA undergoes exo- and endo-nucleolytic cleavages with additional modifications of rRNA by methylation and pseudouridylation to form individual rRNAs (Venema and Tollervey, 1999).

Ribosome biogenesis is one of the essential and highly regulated cellular processes that influence cellular growth and differentiations (Jorgensen *et al.*, 2002). During the last decade, development of various techniques such as affinity purifications followed by mass spectrometry for protein identification has revealed numerous cellular factors including novel genes and RNA components which are involved in ribosome biogenesis and regulation of this process (Takahashi *et al.*, 2003; Merl *et al.*, 2010). Discovery of these factors has enhanced our understanding of the cellular processes of pre-mRNA splicing, mRNA/tRNA turnover, cell cycle progression and, most importantly, the molecular mechanism of the ribosome assembly processes (Fatica and Tollervey, 2002; Fromont-Racine *et al.*, 2003; Tschochner and Hurt, 2003; Kressler *et al.*, 2010).

1.3.2 Translation activation

The activation process consists of the covalent binding of the correct amino acid to the right tRNA. The amino acid is joined by its carboxyl group to the 3'-OH of the tRNA. When the tRNA linked to its corresponding amino acid, it is called charged. Table 1-1 represents possible wobble-pairs at the 1st, 2nd and 3rd bases at two different levels (mRNA and tRNA).

Table 1-1: Possible wobble-pairs at the 1st, 2nd and 3rd bases at mRNA→ tRNA level

1st & 2nd place (5' end)	1st & 2nd place (3' end)
mRNA codon	tRNA anti-codon
A	U
U	A
C	G
G	C
3rd place (3' end)	3rd place (3'end)
mRNA codon	tRNA anti-codon
A or G	U
U	A
U or C	G
G	C
U, C or A	I (Indosine)

1.3.3 Translation initiation

The initiation step is the most complex step of the translation process. Its efficiency often defines the rate of protein synthesis. Initiation of translation is a multistep process that requires many factors and several pre-initiation stages that are mediated by eukaryotic initiation factors (eIF1-eIF6 + PABP) (Pestova *et al.*, 2001; Preiss *et al.*, 2003). Prior to the onset of translation initiation, the eukaryotic 80S ribosome dissociates into 40S and 60S subunits; the dissociated 40S subunit forms a 43S pre-initiative complex in association with eIF1 (essential for transferring the initiator Met-tRNA ternary complex to 40S ribosomal subunit), eIF3 (prevents the large ribosomal subunit from binding the small subunit and is also a platform for the binding of the eIF4F complex), eIF5 (GTPase-activating protein), the Ternary Complex (TC, eIF2-GTP-Met-tRNA, eIF2 is a GTP-binding protein responsible for assigning the initiator tRNA to the P-site of the pre-initiation complex) (Lamphear *et al.*, 1995) and eIF4F complex (composed of eIF4E: cap binding protein, eIF4A: RNA helicase and eIF4G: mRNA-ribosome bridge) which binds to the 5' cap structure (assembly site for 43S complex) in association with eIF4B (Pestova *et al.*, 2001; Preiss *et al.*, 2003). On the other end of the mRNA, the Poly (A)-Binding Protein (PABP) binds to the poly (A) tail at the 3' end. PABP's association with the initiation complex at the 5' end circularizes the mRNA (Sachs and Varani, 2000; Goodfellow and Robert, 2008). Note that eIF6 performs the same ribosomal initiation assembly as eIF3 but binds to the large ribosomal subunit.

The resulting 43S complex starts scanning the mRNA from the cap structure toward the 3' end to locate the first initiation codon (generally an AUG). 48S initiation complex forms during the scanning of mRNA-bound ribosome along the 5'-UTR (43S + AUG).

The initiation codon is base paired to the 48S complex through the anticodon on the tRNA initiator. After recognition of the initiation codon, Met-tRNA binds at the P-site of the ribosome to the initiation codon and forms the AUG-Met-tRNA initiation complex. At this stage, eIF2 and eIF3 are released; this triggers the assembly process of the ribosomal large subunit (60S) and formation of 80S ribosome. At this stage, translation initiation is performed. The translation machinery can now enter its next stage: the elongation phase (Scheper *et al.*, 2007).

1.3.4 Translation elongation

Translation elongation is the stage at which the polypeptide chain is synthesized. During the elongation phase, the ribosome moves along the mRNA from 5' to the 3' direction and sequentially incorporates new amino acids to form a polypeptide chain. The peptidyl-tRNA is translocated into the P-site, leaving the A-site vacant and shifting the ribosome down the mRNA by three nucleotides, positioning a new mRNA codon into the A-site to match with the appropriate tRNA (Kapp and Lorsch, 2004). Unlike the translation initiation, the elongation process is highly conserved between eukaryotes and prokaryotes. In most eukaryotes, elongation process requires two main elongation factors; eukaryotic Elongation Factor 1A (eEF1A) and eukaryotic Elongation Factor 2 (eEF2). eEF1A catalyzes aminoacyl-tRNA binding to the A-site of the ribosome. Once a charged tRNA occupies the A-site, the peptidyl transferase activity of the large subunit promotes a peptide bond between the two amino acids. eEF2 facilitates the translocation stage of peptidyl-tRNA from A-site to the P-site of the ribosome. At this moment, the next codon (triple nucleotide) on the mRNA has moved into the A-site of ribosome. Note that the activity of these proteins is a GTP-dependent process. In addition, yeast requires a third

elongation factor, Elongation Factor 3 (EF3) which is an ATPase (ATP-dependent) and stimulates binding of aminoacyl-tRNA-eEF1A-GTP to the ribosomal A-site by facilitating release of deacylated tRNA from the exit site (Herrere *et al.*, 1984; Sandbaken *et al.*, 1990; Kapp and Lorsch, 2004; Andersen *et al.*, 2006). In agreement with this, an antibody against EF3 has been shown to block natural mRNA translation *in vivo* in yeast (Tariana-Alonso *et al.*, 1995). The translation elongation process is repeated until a stop codon is reached, whereby the termination of translation commences

1.3.5 Translation termination

Translation termination or the termination of protein synthesis occurs when the translation machinery encounters one of the three in-frame termination codons (UAA, UAG and UGA) in the mRNA at the A-site of the ribosome. In eukaryotes, there exist two release factors, eukaryotic Release Factor 1 (eRF1) and eukaryotic Release Factor 3 (eRF3). Stop codons are recognized by eRF1, which then activates the hydrolysis of an ester bond in peptidyl-tRNA to release the polypeptide chain from the ribosome. eRF3 is a GTPase which binds to the ribosome, and facilitates the release of eRF1 (Grentzmann *et al.*, 1998). These two release factors are essential for optimum efficiency of termination process. It has been documented that RFs mimic tRNA and, upon decoding the stop codon, bind to the A-site of the ribosome (Kissele and Buckingham, 2000; Chavatte *et al.*, 2003)). This activates the ribosomal large subunit peptidyl transferase center, which catalyzes the hydrolysis of the peptidyl-tRNA bond and thus releases the polypeptide chain (Kissele and Buckingham, 2000; Chavatte *et al.*, 2003). Once translation has terminated, the ribosome either gets disassembled and its parts recycled for use on a different mRNA or stays assembled to translate the same mRNA again.

1.4 Internal Ribosome Entry Site (IRES)

Besides cap-dependent initiation, translation can be initiated by direct binding of ribosome to specific highly-structured regions of mRNA called Internal Ribosome Entry Sites (IRESs) in association with IRES Trans-acting Factors (ITAFs), Table 1-2, (Le Quesne *et al.*, 2010; Schöler *et al.*, 2006). This is a cap-independent process and it legislates ~10% of the cellular mRNA translation (Stoneley and Willis, 2004). The first IRES was discovered in late 1980s in the *poliovirus* genome. IRESs are often used by viruses to ensure that viral translation is active while host translation process is inhibited. In eukaryotes, it is thought that IRESs facilitate alternative initiation of translation for when the cap-dependent mechanism is blocked under various physiological conditions such as stress (Holcik and Sonenberg, 2005). The cell might also use IRESs to increase the level of translation of certain genes during mitosis or apoptosis; in these processes, cap-dependent translation is blocked or not working properly. For example, in mitosis eIF4F is dephosphorylated and thus has a low affinity for the 5' cap. As a result, the pre-initiation of mRNA loop is not formed so the translation machinery is diverted to the IRES region within the mRNA. In apoptosis (a form of programmed cell death), cleavage (deactivation) of eIF4G (e.g. by viruses) has a negative effect on translation (translation initiation). Proteins essential to perform apoptosis will most likely be produced through IRES-mediated translation. Most natural IRESs are located at 5'UTR (Un-Translated Region) of mRNAs (Holcik and Sonenberg, 2005), but some mediate internal initiation of di/bi-cistronic mRNAs (Hellen and Sarnow, 2001). When an IRES sequence is located between two expression cassettes (reporter gene, ORF) within a eukaryotic mRNA, it might lead to translation of the downstream gene (second gene/ORF) independent of the

5'-capped structure. With this arrangement, both proteins are produced but the first expression cassette (reporter gene) is synthesized through a cap-dependent initiation approach while translation initiation of the second gene is directed by the IRES located in the inter-cistronic spacer region. During IRES-mediated initiation, often a fully assembled ribosome binds to or close to the structural elements of the IRES and scans the mRNA sequence for a nearby initiator codon (Houdebine and Attal, 1999; Komar *et al.*, 2003). Interruptions in IRES detection (IRES-mediated translation) could affect the cap-independent translation pathway.

Table 1-2: Gene Ontology analysis on IRES containing genes (Hellen and Sarnow, 2001; <http://www.iresite.org/> (last updated 2011)). The list of ITAFs continues to grow.

ITAFs gene type	Gene names
Transcription factors	Antennapedia, Ultrabithorax, c-myc, N-myc, MYT2, AMLI/RUNXI, Gtx, c-Jun, Mnt, Nkx6.1, NRF, YAP1, Smad5, HIP-1alpha, Hairless
Stress response factors	XIAP, APC, Apaf-1, Bag-1, Bip/GRP78
Growth factors and growth receptors	FGF2, PDGF2/c-sis, VEGF-A, IGF-II, estrogen receptor a, IGF-1 receptor, Notch2
Translation and RNA processing factors	La, eIF4GI, TIF4631, DAP5/p97/NAT1
Cytoskeletal proteins	ARC, MAP2
Kinases and related	Pim1, p58/PITSLRE, a Cam kinase II, CDK inhibitor p27, PKCd
Channels/transporters	KV-14, bF1-ATPase, Cat-1
Other	Bip, Connexin-43, Connexin-32, Cyr61, ODC, Dendrin, Neurogranin/RC3, NBS1, FMR1, Rbm3, NDST

1.5 Translation process and diseases

The process of protein synthesis is of central importance for the biology of the cell. As it was mentioned before, it is the step at which genetic information becomes meaningful and functional products are produced. During the last two decades, the involvement of the translation pathway (protein synthesis) in numerous neurological diseases (Le Quesne *et al.*, 2010), cellular function, proliferation and tumorigenesis (Clemens, 2004) has been exclusively investigated. It is well documented that cellular abnormal growth requires certain alterations in translation control (gene expression) that clearly makes a significant correlation between cancer and the translation process (Holcik, 2004; Tenesa *et al.*, 2008). Note that many physiological stimuli such as growth factors availability, cell nutrition, and a wide range of stressors such as heat shock, viral infections, DNA damage and exposure to apoptotic condition, have influence on the process of translation and translation regulation (Clemens, 2004).

It is documented that when components of the ribosome are deregulated or mis-expressed, it could lead to onset of various human conditions. For example, mutations in the genes that encode proteins directly involved in ribosome biogenesis, such as the *DKC1* gene, have been linked to dyskeratosis congenital, a disease with premature aging and increased susceptibility to cancer (Ruggero *et al.*, 2003). Another example is the gene encoding the S19 ribosomal protein, whose mutation has been linked to Diamond-Blackfan anaemia, a disease of the bone marrow (Draptchinskaia *et al.*, 1999).

The initiation step of the translation process is the most complex and regulated stage of protein synthesis (Le Quesne *et al.*, 2010). There is a growing body of evidence suggesting the involvement of translation initiation factors in numerous processes such as

regulation of cell cycle, proliferation, transformation and apoptosis. It has been discovered that the overexpression of eukaryotic initiation factor 4E (eIF4E) results in rapid proliferation of cells (West *et al.*, 1995; Anthony *et al.*, 1996; De Benedetti *et al.*, 1999) and might lead into tumor development (Zimmer *et al.*, 2000; Graff *et al.*, 2008). It is also shown that the overexpression of eIF4E and eIF4G also causes oncogenic transformation in different mammalian cell lines (West *et al.*, 1995; Anthony *et al.*, 1996; Raught *et al.*, 1996; Fukuchi-Shimogori *et al.*, 1997). Evidence suggests that the high level of expression of eIF4E-binding protein (eIF4E-BP1) initiates cell growth and promotes apoptosis (Polunovsky *et al.*, 2000). Previous findings suggested that the overexpression of eIF4E might influence human carcinomas including Hodgkin's lymphomas and breast cancer (De Benedetti *et al.*, 1999; Graff *et al.*, 2008). In addition, the overexpression of eIF4G might have an effect on breast (Silvera *et al.*, 2009) and squamous cell lung carcinomas (Brass *et al.*, 1997).

Eukaryotic Initiation Factor 5A (eIF5A) is also reported to be involved in the regulation of cell growth, differentiation and apoptosis. It plays a key role in unusual amino acid hyposine formation that is involved in regulation of cell propagation and transformation (Caraglia *et al.*, 1997).

As previously shown by Schiffmann and Van Der Knaap (2004), vanishing white matter, an autosomal recessive neurodegenerative disease in young children, may be caused by mutation in one of the five subunits of eIF2B.

It is reported that malfunctioning of eEF also plays a role in cytoskeleton regulation (Thornton *et al.*, 2003), cellular growth (Thornton *et al.*, 2003) and neurological diseases (Le Quesne *et al.*, 2010). It has also been reported that in some higher (more complex)

eukaryotes, eukaryotic Elongation Factor I (eEF1) may elevate the rate of protein synthesis (translation efficiency) and causes an increase in missense errors (translation fidelity) (Carr-Schmid *et al.*, 1999). In a recent study, a direct relationship between high levels of eEF1A (eEF1 α) and eEF1B γ (eEF1 γ) and apoptosis acceleration was reported (Duttaroy *et al.*, 1998).

Inhibition of eukaryotic Release Factors (eRFs) (which recognize stop codons at the termination step in translation process) might lead to ribosome stalling which would reduce the rate of protein synthesis within a cell. Previous studies also suggest that eRF3 may control the chromosomal segregation and apoptosis regulations (Malta-Vacas *et al.*, 2005). Therefore, in addition to the important biological function of translation, understanding the role and interactions of factors that make up the protein synthesis machinery is important for understanding human disease.

1.6 Objective

The process of translation has been extensively studied over the past few decades. Although much has been learned about translation, various genome-wide investigations (Krogan *et al.*, 2004; Alamgir *et al.*, 2010; Samanfar *et al.*, 2014) have suggested that there may exist other uncharacterized factors that can affect different stages of the translation process. Interaction and cross-communication of the translation pathway with other cellular processes has long been hypothesized. However, the detailed mechanisms of the factors that affect the translation pathway are in many cases still unknown or unclear. A list of novel genes that affect the process of translation continues to grow. Consequently, we hypothesize that there remains a number of novel genes that can affect

the process of translation in yeast. Here, we aim to identify some of these novel genes that can affect the translation pathway. *S. cerevisiae* is used as a model organism for our investigation. Analyzing the activity of these genes can increase our understanding of the process of translation and its communication with other cellular processes. Our specific objectives are as follows:

1. Identification and investigation of novel genes involved in stop codon read-through (an aspect of translation fidelity) in yeast using yeast non-essential knockout collection.
2. Utilization of the yeast genome to identify novel genes that affect IRES-mediated translation (IRES recognition).
3. Utilizing a computational approach to identify novel candidates involved in the translation pathway on the basis of PPI predictions.
4. Yeast genome-wide investigation to identify novel players in translation that can affect the cells abilities to unwind mRNA inhibitory structures.

2 Chapter: A global investigation of gene deletion strains that affect premature stop codon bypass in yeast, *Saccharomyces cerevisiae*

2.1 Abstract

Protein biosynthesis is an orderly process that requires a balance between rate and accuracy. To produce a functional product, the fidelity of this process has to be maintained from start to finish. In order to systematically identify genes that affect stop codon bypass, three expression plasmids pUKC817, pUKC818 and pUKC819, were integrated into the yeast non-essential loss-of-function gene array (~5000 strains). These plasmids contain three different premature stop codons (UAA, UGA and UAG, respectively) within the *LacZ* expression cassette. A fourth plasmid, pUKC815 that carries the native *LacZ* gene was used as a control. Transformed strains were subjected to large-scale β -galactosidase lift assay analysis to evaluate production of *β -galactosidase* for each gene deletion strain. In this way 84 potential candidate genes that affect stop codon bypass were identified. Three candidate genes, *OLA1*, *BSC2*, and *YNL040W*, were further investigated, and were found to be important for cytoplasmic protein biosynthesis.

2.2 Introduction

Protein biosynthesis has been extensively studied over the past few decades. Although much has been learned, details of the genes that regulate and influence this process require further investigation. The list of new genes that affect protein biosynthesis continues to grow (Alamgir *et al.*, 2010; Galvao *et al.*, 2013; Hopper, 2013). Recent large-scale investigations suggest that there exist other uncharacterized factors that

influence protein biosynthesis (Alamgir *et al.*, 2010; Galvao *et al.*, 2013; Hopper, 2013; Tafforeau *et al.*, 2013)

To produce functional proteins, the accuracy of protein biosynthesis is tightly regulated from start to finish. Errors in start codon selection during initiation, misincorporation of amino acids during elongation or inability of the ribosome to detect a stop codon during translation termination can alter the integrity of the protein biosynthesis process.

Efficient termination is an important step in ensuring accuracy of the synthesized products (Janzen and Geballe, 2001; Ehrenreich *et al.*, 2012). A key determinant of the precision of translation termination is the ability of release factors to correctly recognize the stop codons. Reduced cellular levels of eukaryotic release factor 1 (eRF1) has been linked to higher stop codon bypass (Stansfield *et al.*, 1996). Similarly premature stop codon mutants of *SUP45* which codes for eRF1, are reported to exhibit increased frequency of stop codon bypass (Moskalenko *et al.*, 2003). Nonsense suppressor tRNAs can also increase the bypass of stop codons by incorporating amino acids into a growing polypeptide chain when a stop codon is reached (Gubbens *et al.*, 2010).

In the current study, we used the yeast non-essential gene knockout (also known as loss-of-function) collection to identify genes that affect stop codon bypass. To this end, 84 candidate genes were identified that once deleted, increased the synthesis of full length protein directed from an mRNA with premature stop codon. Many of these genes have unknown functions. We further studied the activity of three of these genes, *OLAI*, *BSC2* and *YNL040W*.

2.3 Materials and Methods

2.3.1 Strains

The Yeast deletion and overexpression set (*Saccharomyces cerevisiae*, mating type “a” (BY4741)) were used for large-scale investigation. Yeast mating type “α” (BY7092) was used for secondary gene knockout (Tong *et al.*, 2001). DH5α strain of *Escherichia coli* was used to propagate plasmids (Taylor *et al.*, 1993).

2.3.2 Media and antibiotics

Standard rich (YPD) and synthetic complete (SC) media were used as growth media for yeast. LB (Lysogeny broth) was used to growth *E.coli*. Antibiotics were used in the following concentrations: Paromomycin (18 mg/ml) and Cycloheximide (45 ng/ml) for drug sensitivity analysis; G418, Geneticin (200 mg/ml), ClonNat, Nourseothricin, (100mg/ml) and Ampicillin (50 mg/ml) for selective growth media (Tong *et al.*, 2001; Alamgir *et al.*, 2008, 2010).

2.3.3 Plasmids

Expression plasmids pUKC817, pUKC818 and pUKC819, contain premature stop codons UAA, UGA and UAG, respectively, within the LacZ expression cassette; pUKC815 has no premature stop codon and was used as a control (Lucchini *et al.*, 1984; Alamgir *et al.*, 2008). These plasmids were used to transform in to the yeast non-essential gene knockout collection using a modified synthetic genetic array analysis approaches (Tong *et al.*, 2001; Alamgir *et al.*, 2008). Translation rate was assessed using plasmids p416 (Gal-inducible LacZ expression cassette) and pAM6 (CoCl₂-inducible LacZ expression cassette) as described by (Vasconcelles *et al.*, 2001; Krogan *et al.*, 2003;

Alamgir *et al.*, 2008). pAG25 was used as a template in PCR reactions to propagate NAT resistance gene.

2.3.4 Gene knockout

Gene knockout was performed via homologous recombination by using the LiAc-based transformation method described by (Inoue *et al.*, 1990). In this technique, the gene of interest is removed and replaced with a NAT antibiotic resistant marker. Knockout strains were confirmed using colony PCR analysis (appendix 2-1) and were used for small-scale follow-up experiments.

2.3.5 Primers

All primers used in this study have been described in appendix 2-1.

2.3.6 qRT-PCR

High quality extracted RNA was converted to cDNA using iScript cDNA synthesis kit (Biorad) according to manufacturer's instruction. Quantitative PCR was performed using iQSybergreen master-mix kit (Biorad) according to manufacturer's instruction. qPCR amplification and detection was performed on a Rotor Gene 3000 (Corbett Research). Thermo cycler conditions were set to the following: 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 30s-60°C for 30s-72°C for 30s and a final 72°C for 10 min (Pfaffl *et al.*, 2001; Yu *et al.*, 2007b). PGK-1 is used as a positive control in real time PCR experiments (Chambers *et al.*, 1989).

2.3.7 Genetic material extractions

Total RNA was extracted using RNeasy Mini Kit (QIAGEN) according to manufacturer's instruction. Plasmids were extracted from *E.coli* using pure link quick plasmid kit

(Invitrogen) and from yeast using yeast plasmid kit (Omega Bio-Tek) according to manufacturer's instruction.

2.3.8 *β*-galactosidase assay

A large-scale *β*-galactosidase assay using X-gal was carried out as described (Favier *et al.*, 1996, 1997; Serebriiskii and Golemis, 2000). The final incubation step was continued until an average of ~5 colonies per plate (approximately 384 colonies) turned blue (approximately 45 minutes). Quantitative *β*-galactosidase assay was performed using ONPG (O-nitrophenyl- α -D-galactopyranoside) as described by (Lucchini *et al.*, 1984; stansfield, 1995). The rate of protein biosynthesis was measured using inducible reporter systems p416 and pAM6 (Vasconcelles, 2001; Alamgir *et al.*, 2008). Yeast strains transformed with expression plasmids were grown to OD₆₀₀: 0.8-1 and induced by addition of either 2% galactose (p416) or 40 μ M cobalt chloride (pAM6) for 4 hours. *β*-galactosidase activity was then quantified.

2.3.9 Spot test (drug sensitivity analysis)

Yeast cells were grown in liquid media to mid-log phase and diluted to an OD₆₀₀ of 0.01, 15 μ l of this suspension and three subsequent 10 fold dilutions were plated onto solid media containing translation inhibitory drugs (paromomycin and cycloheximide). Media with no antibiotics was used as a control. All plates were incubated at 30°C for 1-2 days in 3 replications.

2.3.10 Synthetic genetic array (SGA) and phenotypic suppression array (PSA) analyses

Synthetic genetic array (SGA) analysis was performed on the basis of yeast haploid double gene deletion mutant growth defects. Colony size is measured as the basis for a

growth deficiency. The query genes OLA1, BSC2 and YNL040W were replaced with the nourseothricin-resistance (NAT) marker forming gene deletion strains in the MAT α strain, BY7092. The deletion strains were separately crossed to two target sets of deletion mutant arrays from the MAT α BY4741 library. One of the sets contained gene deletion mutants for protein biosynthesis genes (selected on the basis of GO terms) and the other contained deletion mutant for random genes, minus those in set one, and used as a control. After a few rounds of selection using G418 + NAT selective media, haploid double mutant progeny were obtained and analyzed for growth defects using colony size measurements (Memarian *et al.*, 2007; Samanfar *et al.*, 2013). The experiment was repeated three times and those interactions that were found in at least two experiments were considered for confirmation using random spore analysis (Tong *et al.*, 2001). Phenotypic suppression array (PSA) analysis was performed as described by (Sopko *et al.*, 2006; Alamgir *et al.*, 2010).

2.4 Results and Discussion

2.4.1 Identification of gene deletions that affect expression of β -galactosidase gene with premature stop codons

In order to identify novel genes involved in stop codon bypass (read-through) three plasmids, pUKC817, pUKC818 and pUKC819 that carry different premature stop codons, UAA, UGA and UAG, respectively within a β -galactosidase reporter gene were used to transform into the yeast non-essential gene deletion array (~5000 strains). The plasmid pUKC815 which carries the native β -galactosidase gene (used as a control) was also transformed for a total of approximately 20,000 strain transformations. Systematic transformation was accomplished by the modification of a large-scale mating approach

that was originally designed to study high throughput genetic interactions in yeast (Tong *et al.*, 2001). In this method, a query yeast strain of “ α ” mating type carrying a selectable marker (in this case URA selection derived from a plasmid) is crossed with the gene deletion array strains of “ a ” mating type. After several rounds of selection, gene deletion strains of “ a ” mating type carrying the selectable marker of the query strain are selected. In this way, an array of strains each carrying a specific gene deletion together with a target plasmid is generated. The arrayed colonies are tested for their ability to produce functional β -galactosidase using a large-scale colony lift assay on the basis of X-gal hydrolysis. The premature stop codon within the β -galactosidase gene would prevent the production of full length functional β -galactosidase. Under high-fidelity, the premature stop codon is recognized, mediating the synthesis of a short non-functional polypeptide. Only when the premature stop codon is bypassed, a full length β -galactosidase protein is produced. Deletion of genes that affect premature stop codon recognition would allow the bypass of stop codons and result in increased production of full length β -galactosidase. Figure 2-1 illustrates a representative β -galactosidase array analysis where blue colonies are those that produced higher levels of functional β -galactosidase, recognizable by visual inspection. Each lift assay was repeated 3 times and those colonies that were detected in 2 or 3 experiments were selected.

Figure 2-1: A representative yeast gene deletion colony array that was subjected to β -galactosidase lift assay. Blue colonies indicate the presence of higher levels of β -galactosidase produced from an mRNA that carries a premature stop codon.

In this way 84 gene deletion strains were identified that produced blue color for one or more of the plasmids with premature stop codons.

To confirm that our large-scale screening correctly identified those colonies that produced functional *β-galactosidase*, we alphabetically ordered and selected the first, tenth, twentieth, thirtieth, ... deletion strains (9 in total) to follow up analysis using a small scale (low-throughput) quantitative liquid based assay on the basis of ONPG hydrolysis. We observed that 7 of 9 gene deletion strains produced levels of functional *β-galactosidase*, which were statistically higher than that produced by a control strain. This suggests an overall sensitivity of approximately 78% for our screen. The list of the genes that were identified in our screen is found in Appendix 2-2. Among these we find a number of expected genes such as *RPL31A* that codes for large ribosomal subunit protein L31A, *NAM7* that codes for Upf1 protein required for efficient termination at nonsense codons, *NOP58* that codes for Nop58 protein involved in pre-rRNA processing, etc.

We subjected the gene deletion strains identified via our large-scale screening to sensitivity analysis against two antibiotics, cycloheximide and paromomycin. Cycloheximide interferes with the translocation step of protein biosynthesis (Obrig *et al.*, 1971; Alamgir *et al.*, 2010) and paromomycin has been linked to codon misreading (Tuite and McLaughlin, 1984). Using serial dilution spot test analysis, decreasing cell concentration of the 84 target strains were subjected to sub-inhibitory concentrations of drugs; 45 ng/ml of cycloheximide and 18 mg/ml of paromomycin. Approximately 69% of the identified gene deletion strains showed sensitivity to one or more drugs (appendix 2-2). This compares to approximately ~26% of the entire set of yeast non-essential gene

deletion strains that show sensitivity to at least one of these drugs (Alamgir *et al.*, 2008, 2010).

To improve understanding of the activity of the identified genes, we performed Gene Ontology (GO) enrichment analysis to cluster these genes on the basis of the function or the cellular process in which they participate. GO analysis indicates that, 19% of our selected genes are involved in protein biosynthesis pathways (P-value: $2.40\text{E-}8$), 12% are involved in regulation of nucleic acids synthesis/stability (P-value: $3.51\text{E-}02$), 7% are involved in cell wall/membrane synthesis (P-value: $3.33\text{E-}02$), 6% are involved in sulfate synthesis (P-value: $5.72\text{E-}06$) and 5% are involved in signal transduction (P-value: $1.07\text{E-}02$). In addition 31% of the genes were found to have either unknown function and/or unknown cellular process in which they participate. Some of the clustering observed here might be expected. It is not surprising to detect protein biosynthesis related genes. Mutations in these genes can influence the overall process of translation, and hence the ability of the cell to correctly recognize different codons. Similarly, deficiency in regulation of nucleotide stability through nonsense-mediated mRNA decay could explain an increase in protein expression from an mRNA template with premature stop codon, by preventing the mRNA from degradation. The less expected categories are cell wall/membrane synthesis, sulfate synthesis and signal transduction. A cross-communication between protein biosynthesis and cell wall/membrane might be explained by a proposed regulation of translation initiation fidelity which is mediated by stress (Nierras and Warner, 1999). Similarly, very recently cell wall integrity has been connected to protein biosynthesis through the activity of eukaryotic translation initiation factor 5A (*eIF5A*) (Galvao, 2013). Sulfur modification of translation machinery has been

linked to fidelity of translation (Hopper, 2013) and hence can support an enrichment of sulfate pathway genes among those that affect stop codon bypass. Lastly, enrichment for signal transduction genes may be explained by the activity of signal transduction pathway in the control of eukaryotic protein biosynthesis (Rhoads, 1999).

Little information is known about the remaining 31% of the genes. The list of these includes *OLAI*, a conserved gene whose product is homologous to a sub-family of A(G)TPases linked to translation in human cells (Eichhorn *et al.*, 2007), *BSC2*, an uncharacterized gene that was previously found to contain a region (± 50 nucleotides surrounding the stop codon) that promotes stop codon bypass (Namy *et al.*, 2003) and *YNL040W*, an uncharacterized open reading frame. Knockout mutant strains for *OLAI*, *BSC2* and *YNL040W* genes showed sensitivity to both cycloheximide and paromomycin (Figure 2-2). Re-introduction of these genes into their corresponding gene knockout strains reversed their sensitive phenotypes (Figure 2-2). We further investigated the activity of these genes.

2.4.2 *OLAI*, *BSC2* and *YNL040W* affect protein biosynthesis

Observed influence of gene deletions for *OLAI*, *BSC2* and *YNL040W* on premature stop codon bypass was confirmed using low-throughput β -galactosidase assay on the basis of ONPG hydrolysis. As indicated in Figure 2-3A, deletion of each of the target genes increased the level of β -galactosidase expression derived from all three expression cassettes carrying different premature stop codons. The observed differences were comparable to those for *rpl31aA* strain used as a positive control. Alteration in mRNA content can also explain differences in gene expression. To examine this possibility, the content of β -galactosidase mRNAs for the gene deletion strains was investigated using

qRT-PCR analysis. No statistically significant difference between the content of mRNAs for the target and control strains were observed (Figure 2-3B). This indicates that the observed increase in expression of *β-galactosidase* mRNAs with premature stop codons appears to be at the protein biosynthesis level and not at the mRNA content level.

Integrity of protein biosynthesis is maintained through a balance between the accuracy and the rate of translation. Next we asked whether deletions of *OLA*, *BSC2* and *YNL040W* had any effect on the rate of translation. This was evaluated using two inducible *β-galactosidase* expression systems (p416 and pAM6). It was observed that deletion of *YNL040W* significantly reduced the rate of protein biosynthesis measured by *β-galactosidase* expression (Figure 2-4A). Interestingly, however, deletion of *OLA1* and *BSC2* increased the level of protein biosynthesis (*rpl31aΔ* was used as a positive control). As a possible source of alteration of gene expression, the content of *β-galactosidase* mRNA was investigated as above. The content of *β-galactosidase* mRNA was found to be constant in all gene deletion strains suggesting that the observed differences appear to be independent of mRNA content (Figure 2-4B).

Figure 2-2: Drug sensitivity analysis. Gene knockout mutant strains for *OLA1*, *BSC2* and *YNL040W* showed sensitivity to both cycloheximide and paromomycin. Re-introductions of the deleted genes into the corresponding gene knockout mutants reversed the observed sensitivity. Yeast cells were grown to mid-log phase and diluted to an OD₆₀₀ of 0.01, 15µl of this suspension and three subsequent 10 fold dilutions were plated.

Figure 2-3: A) β -galactosidase expression from templates that contain premature stop codons. Constructs pUKC817, pUKC818 and pUKC819 contain premature stop codon UAA, UGA and UAG, respectively, within *LacZ* expression cassette. pUKC815 contains native *LacZ* gene. β -galactosidase expression derived from constructs with premature stop codons (pUKC817, pUKC818 and pUKC819) are normalized to β -galactosidase expression from a native *LacZ* gene (pUKC815) and related to that of a control strain. B) β -galactosidase mRNA content analysis. β -galactosidase mRNA contents are related to those for the control strain. *PGK1* mRNA content was used for normalization*(P-value $\leq$ 0.05).

Figure 2-4: A) Translation rate measurement. Translation rate was measured using two β -galactosidase reporter expression cassettes, p416 and pAM6, under the transcriptional control of inducible Gal1 and LORE promoters, respectively. The values are normalized to that of a control strain. B) β -galactosidase mRNA content analysis. β -galactosidase mRNA content was related to that of *PGK1* for the above inducible β -galactosidase constructs. *(P-value $\leq$ 0.05).

Considering that deletions of *OLAI*, *BSC2* and *YNL040W* affected the stop codon bypass (above), one way to explain these results is that *OLAI* and *BSC2* may participate in maintaining a balance between the quantity and quality of protein biosynthesis. Deletions of *OLAI* and *BSC2* increased the rate of protein biosynthesis, but decreased the ability of the cell to recognize stop codons. The effect of *YNL040W* deletion seems to be more general, reducing both the rate of translation and recognition of premature stop codons.

To better investigate the involvement of *OLAI*, *BSC2* and *YNL040W* in protein biosynthesis we, studied the genetic interactions they made with other genes that influence translation. It is possible that alterations in the expression of two genes result in a phenotype that is not easily explained in light of the phenotypes caused by alteration of individual gene expressions. In this case, the two genes are said to form a genetic interaction (Tong *et al.*, 2001; Boone *et al.*, 2007). Genetic interactions can be used to investigate functional association between two genes. The most commonly studied form of genetic association is explained by negative genetic interactions (Tong *et al.*, 2001; Alamgir *et al.*, 2008; Samanfar *et al.*, 2013). In a given pathway “A”, deletion of a target gene “A1” may have a negligible phenotypic consequence due to the presence of a parallel pathway “B” that compensates for the absence of “A1” gene product. Similarly, deletion of a gene in pathway “B” can be compensated by pathway “A”. However, deletion of two genes, one in pathway A (for example, gene “A1”) and the other in pathway B (for example, gene “B2”), can leave both pathways inactive and hence cause a phenotypic consequence (synthetic sick phenotype) that cannot be explained by the phenotypes for individual gene deletions for genes “A1” and “B2” alone. In this way, genes “A1” and “B2” are described to form a negative genetic interaction. These

interactions can provide insight into a higher level association between gene functions. In this manner, function(s) of uncharacterized genes may be studied by the genetic interactions that they make with other genes with known functions (Hsu *et al.*, 2001; Alamgir *et al.*, 2008; Samanfar *et al.*, 2013). To this end, we investigated the synthetic sick interaction that *OLAI*, *BSC2* and *YNL040W* formed with two sets of 384 gene deletion strains. Set one contained an array of deletion mutant strains for genes associated with the process of protein biosynthesis and set two contained strains with deletion mutations for random genes (excluding those involved in protein biosynthesis), used as a control. To study negative genetic interactions, gene deletion strains for target genes were generated in an “ α ” mating type strain. Then, the query strains were crossed with the array of gene deletion strains of “a” mating type. After a few rounds of selection, double gene deletion mutants in “a” mating type were selected. In this way 768 double gene deletion mutants were generated for each target gene. The fitness of double gene deletion strains were quantified and analyzed by colony size measurements (Memarian *et al.*, 2007; Samanfar *et al.*, 2013). To improve coverage, we combined our interaction data with those previously reported (<http://drygin.ccbr.utoronto.ca>). A complete list of interactions is presented in appendix 2-3. Illustrated in Figure 2-5A, *OLAI* formed negative interactions with a number of interesting genes including ribosomal protein encoding genes (for example, *RPL12A* and *RPL15B*), *GCN1* that codes for a positive regulator of translation initiation factor eIF2 by Gcn2, and *EFT1* that codes for elongation factor 2 involved in ribosomal translocation. Further clustering of the interacting genes into different functional categories on the basis of GO terms showed enrichment for 4 clusters. The most significant enrichment belonged to genes involved in regulation of

translation (P-value: 2.27E-07). This was followed by ribosome biogenesis (P-value: 4.37E-05), amino acid metabolism (P-value: 1.26E-03) and RNA binding (P-value: 5.11E-03). Similarly, *BSC2* interacted with a number of protein biosynthesis related genes; clustering of the interacting partners showed enrichment for genes involved in regulation of translation (P-value: 8.94E-04) (Figure 2-5B). *YNL040W*s interactors were enriched for ribosome biogenesis gene (P-value: 1.21E-02) (Figure 2-5C). Re-introduction of *OLA1*, *BSC2* and *YNL040W* into a representative set of corresponding double mutant strains, reversed the sick phenotype observed for the double mutant strains (appendix 2-4).

Next we examined the ability of the overexpression of *OLA1*, *BSC2* and *YNL040W* genes to compensate the phenotypes of different gene deletion mutants in response to cycloheximide treatment. For this, phenotypic suppression analysis (PSA) was used. This is a similar approach to the genetic interaction analysis method above with the exception that a compensatory effect of the overexpression of a target gene is sought. If the overexpression of a target gene compensates a phenotype caused by the absence of another gene, a functional link between the two genes is assumed (Sopko *et al.*, 2006; Alamgir *et al.*, 2008, 2010). To this end, the strains in the two gene deletion arrays described above (one for protein biosynthesis related genes and the other random) were separately transformed with overexpression plasmids for *OLA1*, *BSC2*, and *YNL040W*, in addition to an empty plasmid. Gene deletion strains along with those transformed with overexpression plasmids were grown in the presence of a sub-inhibitory concentration of cycloheximide (45 ng/ml).

Figure 2-5: The union (from this study and literature) of synthetic sick interactions for *OLA1* (A), *BSC2* (B) and *YNL040W* (C). *OLA1* interacted with genes that contain the following GO terms: regulation of translation (P-value: 2.27E-07), ribosomal biogenesis (P-value: 4.37E-05), amino acid metabolism (P-value: 1.26E-03) and RNA binding (P-value: 5.11E-03). B) *BSC2* interacted with gene involved in regulation of translation (P-value: 8.94E-04). C) *YNL040W* interacted with ribosomal biogenesis genes (P-value: 1.21E-02). Degree of sickness is color coded. * Represents interactions that were included from literature.

Those gene deletion strains that showed sensitivity to the drug treatment, but whose sensitivity was compensated by one of the target overexpressed genes, were identified on the basis of colony size measurement. Positive hits were confirmed using spot test analysis (Figure 2-6). It was observed that overexpression of *OLAI* gene compensated for sensitivity to cycloheximide caused by deletions of *RPS11B*, *RPS8A* and *CPA1* genes. *RPS11B* and *RPS8A* code for small ribosomal subunit proteins S11 and S8, respectively. S11 forms part of the mRNA exit tunnel and S8 is part of a bridge between 40S and 60S subunits (Shem *et al.*, 2011). *CPA1* codes for carbamoyl phosphate synthetase involved in arginine biosynthesis. *BSC2* overexpression compensated for gene deletion for *RPS24B* which codes for small ribosomal subunit S24B. Overexpression of *YNL040W* compensated for the deletions of *DEP1* and *SIN3*, both of which code for components of Rpd3L histone deacetylase complex shown to affect the expression of rDNA genes (Johnson *et al.*, 2013). The phenotypic compensations observed here further connect the activity of the target genes to protein biosynthesis.

Figure 2-6: Representative spot test confirmation for phenotypic suppression analysis of *OLA1*. Gene deletion mutants for *RPS11B*, *RPS8A* and *CPA1A* show sensitivity to 45 ng/ml of cycloheximide. Overexpression of *OLA1* compensates the observed sensitivity. Yeast cells were grown to mid-log phase and diluted to an OD₆₀₀ of 0.01, 15μl of this suspension and three subsequent 10 fold dilutions were plated.

Little information is available about the activity of yeast Ola1 protein. Here, we show that deletion of *OLA1* can increase premature stop codon bypass. It also increased the rate of protein biosynthesis measured by an inducible expression system. *OLA1* formed negative genetic interactions with a number of protein biosynthesis related genes most notably those involved in translation regulation. Its overexpression also rescued phenotypes for *RPS11B*, *RPS8A* and *CPA1*, further connecting Ola1 activity to protein biosynthesis. An involvement for Ola1 in protein biosynthesis is supported by its expression data from microarray analysis (Wyrick *et al.*, 1999; Brem and Kruglyak, 2005). The group of proteins with the most similar profile of expression to Ola1 protein are highly enriched for cytoplasmic translation genes (P-value: 1.29E-8). These observations are in agreement with what is known about human Ola1 (hOla1) protein. It is characterized as a member of obg family comprising a group of ancient A(G)TPases that belong to translation factor (TRAFAC) proteins (Eichhorn *et al.*, 2007).

Bsc2 is another protein of unknown function. Its genomic region is reported to contain a section with a premature stop codon compatible with the Bypass of Stop Codon (BSC) mode of expression (Namy *et al.*, 2003). Here, we report that its deletion increased the rate of premature stop codon bypass. Consequently, it appears that the Bsc2p may function in negative regulation of premature stop codon bypass and hence regulate its own expression. To our knowledge this is a unique mode of regulation of eukaryotic gene expression and demands further investigation. Interestingly, deletion of Bsc2 also increased the rate of protein synthesis further connecting the activity of Bsc2 to regulation of translation. This is further supported by negative genetic interaction data where *BSC2* interacted with a number of genes involved in regulation of translation.

YNL040W codes for a putative protein of unknown function with similarity to bacterial alanyl-tRNA synthetase. Here, we show that deletion of *YNL040W* increased the rate of premature stop codon bypass and reduced the rate of protein synthesis connecting its activity to the process of protein biosynthesis. The genetic interaction data further links *YNL040W* to ribosome biogenesis.

3 Chapter: The identification and investigation of 7 novel genes involved in IRES-mediated translation in yeast, *Saccharomyces cerevisiae*

3.1 Abstract

Translation often to the multifaceted processes that together lead to the synthesis of protein products from genetic information. In eukaryotes, beside cap-dependent translation, protein synthesis can be initiated by the means of an internal sequence within mRNAs. This alternative initiation is cap-independent and is called Internal Ribosome Entry Site (IRES) mediated translation initiation. *HSP82* and *HSC82* code for two yeast heat shock proteins, deletion of which alter sensitivity to acetic acid treatment. Hsp82 and Hsc82 are shown to be translationally regulated via an IRES-mediated mode of translation initiation. In the current study we identify 7 additional gene deletion mutants (*att22Δ*, *mrps16Δ*, *pcs60Δ*, *rpl36aΔ*, *vps70Δ*, *ybr063cΔ* and *dom34Δ*) that show sensitivity to acetic acid. We investigated the influence of these gene deletions on the expression of Hsp82 and Hsc82. Our results indicate that the target genes affected the expression of Hsp82 and Hsc82 at the translation level, influencing IRES-mediated translation initiation.

3.2 Introduction

In general, cells respond to stress in different manners ranging from production of by-products to programmed cell death (PCD). The *in-vivo* effect of various stressors including hydrogen peroxide, acetic acid and heat shock have been investigated in depth

using *Saccharomyces cerevisiae* budding yeast as a model system (Madeo *et al.*, 1999; Ludovico *et al.*, 2001; Ludovico and Madeo, 2005; Silva *et al.*, 2013). In this context acetic acid is of particular interest as it is a by-product produced by various microorganisms including yeast itself through the fermentation process. Acetic acid has been identified to affect cell viability and cause apoptosis/PCD (Ludovico *et al.*, 2001; Silva *et al.*, 2013).

Translational control is a key regulatory step in regulation of gene expression. Generally speaking, translation control can be divided into two main modes, a global control (general) and an mRNA specific control (selective). Structural features and regulatory sequences found in the mRNAs including inhibitory mRNA structures and IRES are examples of regulatory elements that can control selective translation. In eukaryotes, the general mode of translation initiation is thought to be dependent on the 5'-cap scanning mechanism. However, in an alternative mechanism, ribosomes and associated factors called IRES Trans Acting Factors (ITAF) can bind to mRNA directly in the vicinity of the start codon. Such internal interactions were originally discovered in viral systems and explain how invading genome is expressed when infected cells' translation systems shut down (Hellen and Sarnow, 2001; Komar *et al.*, 2003; Filbin and Kieft, 2009; King *et al.*, 2010; Komar and Hatzoglou, 2011; Thakor and Holcik, 2012). In eukaryotes including yeast, it is thought that canonical, cap-dependent translation is generally responsible for proteins required by the cell in bulk amounts, while IRES is responsible for certain essential proteins required in smaller quantities (Spriggs *et al.*, 2008; Thompson, 2012). This constitutes approximately 10-15% of cellular mRNAs.

Multiple mutational studies have been performed on IRES regions to identify sequences crucial to initiation activity (Andreev *et al.*, 2009; Zhu *et al.*, 2011; Thakor and Holcik, 2012); although this method was successful in abolishing activity of specific transcripts, no particular consensus sequence has been identified. Recent studies in mammalian systems have shown that IRES-mediated translation is implicated in certain genes that control cellular stress response, life cycle progression and cell survival (Allam and Ali, 2010; Komar and Hatzoglou, 2011; Thakor and Holcik, 2012). These include *c-myc*, Apaf-1, p53, XIAP, VEGF, HSP90, and c-Src all of which some are linked to cell cycle control and implicated in cancer (Allam and Ali, 2010; Komar and Hatzoglou, 2011; Silva *et al.*, 2013). One of these genes, Hsp90, is a highly abundant and conserved molecular chaperone which plays a central role in various cellular processes including cell cycle control, cell survival, signal transduction, intracellular transport, and protein degradation (Jackson, 2013). Hsp90 has two major isoforms: Hsp90 α which is inducible under stress and Hsp90 β which is constitutively expressed (Langer *et al.*, 2003; Ahmed and Duncan, 2004). A selective translation mechanism is implicated to be part of HSP90 induction (Ahmed and Duncan, 2004). Mammalian Hsp90 undergoes selective translation that regulates cellular response to stress (Langer *et al.*, 2003). However, the precise control mechanism of HSP90 translation is yet to be understood. In yeast, there are two Hsp90 homologs, known as Hsc82 and Hsp82, of which Hsp82 is known to be up-regulated during heat shock stress (Borkvich *et al.*, 1989). In this study, we have identified 7 yeast gene deletion mutants (*att22 Δ* , *mrps16 Δ* , *pcs60 Δ* , *rpl36a Δ* , *vps70 Δ* , *ybr063c Δ* and *dom34 Δ*) that show sensitivity to acetic acid treatment. We further investigated their activity on the expression of Hsp82 and Hsc82 in yeast.

3.3 Material and Methods

3.3.1 Yeast strains, media and plasmids

Yeast, *Saccharomyces cerevisiae*, strains are picked from the library of gene-deletion mutants derived from the MATa strain BY4741 (*MATa orfΔ::KanMAX4 his3Δ1 leu2Δ0 met15Δ0 ure3Δ0*) (Winzeler *et al.*, 1999; Tong *et al.*, 2001) and the MATa strain, BY7092 (*MATa Can1Δ::STE2pr-HIS3 Lyp11Δ leu21Δ0 his31Δ0 met151Δ0 ure31Δ0*) (Tong *et al.*, 2001). YPD (Yeast extract, Peptone, Dextrose) and SC (Synthetic Complete) for yeast and LB (Lysogeny broth) for *Escherichia coli* were used as growth media. Expression plasmids p281-4-HSC82, p281-4-HSP82, p-281-4-URE2, contain IRES sequences, respectively, fused into the upstream of LacZ expression cassette; p281-4-CLN3 with no IRES sequences was used as a negative control (Silva *et al.*, 2013; Komar *et al.*, 2003). pAG25 was used as a DNA template in PCR reactions to propagate NAT resistance marker gene for secondary gene knockout.

3.3.2 Gene knockout and DNA transformation

Gene knockout was performed using homologous recombination *via* LiAc-based transformation method described by (Inoue *et al.*, 1990). Knockout strains were confirmed via colony PCR.

3.3.3 Primers

All primers used in this study have been described in Appendix 3-1.

3.3.4 Genetic material extractions

Total RNA was extracted using RNeasy Mini Kit (QIAGEN) according to manufacturer's instruction. Plasmids were extracted from *E.coli* and yeast using pure link quick plasmid kit (Invitrogen) according to manufacturer's instruction.

3.3.5 qRT-PCR

High quality extracted RNA were converted into cDNA using iscript cDNA synthesis kit (Biorad) according to manufacturer's instruction. Quantitative PCR was performed using iQSybergreen master-mix kit (Biorad) according to manufacturer's instruction. qPCR amplification and detection was performed on a Rotor Gene 3000 (Corbett Research). Thermo cycler conditions were set to the following: 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 30s-60°C for 30s-72°C for 30s and a final 72°C for 10 min (Pfaffl *et al.*, 2001; Yu *et al.*, 2007b). PGK1 was used as a positive control, housekeeping gene, in real time PCR experiments (Chambers *et al.*, 1989; Samanfar *et al.*, 2013; 2014).

3.3.6 β -galactosidase assay

A quantitative, plasmid-based β -galactosidase assay was performed using the substrate ONPG (O-nitrophenyl-beta-D-galactopyranoside) as described by (Lucchini *et al.*, 1984; Stansfield *et al.*, 1995).

3.3.7 Spot test (Drug sensitivity analysis)

A series of single and double deletion strains (mutants) grown to the mid-log phase were diluted (10^{-2} - 10^{-5}); 15 μ l of each dilution were plated onto solid media containing acetic Acid (220 mM for 2 hours of treatment). Media with no drug was used as a control. All plates were incubated at 30°C for 1-2 days in 3 replications.

3.3.8 Synthetic Genetic Array (SGA) and Phenotypic Suppression Array (PSA) analysis

Synthetic genetic array (SGA) analysis was performed on the basis of yeast haploid double mutant growth defects (sickness or lethality). Colony sizes were measured as the

basis for a growth deficiency. The query genes (*ATP22*, *MRPS16*, *PCS60*, *RPL36A*, *VPS70*, *YBR063C* and *DOM34*) were deleted and replaced with the nourseothricin-resistance (NAT) marker forming gene deletion strains in the MAT α strain, BY7092. The deletion strains were separately crossed into two target sets of deletion mutant arrays from the MAT α library (single gene deletion replaced with KanMAX resistance marker) (BY4741). One of the sets (called a translation array) contained gene deletion mutants for protein biosynthesis genes (selected on the basis of GO terms) and the other contained deletion mutant for random genes, minus those in set one, and was used as a control (a total of 768 mutants). After a few rounds of selection using G418 + NAT selective media, haploid double mutant offspring were obtained and analyzed for growth defects using colony size measurements (Memarjal *et al.*, 2007; Samanfar *et al.*, 2013, 2014). The experiment was repeated in triplicate and those interactions that were found in at least two experiments were considered for confirmation using random spore analysis (Tong *et al.*, 2001). Phenotypic suppression array (PSA) analysis was performed as described by (Sopko *et al.*, 2006, Alamgir *et al.*, 2008; Samanfar *et al.*, 2014).

3.4 Results and Discussion

3.4.1 Identification of novel candidate genes affect IRES-mediated translation pathway

Mechanistically, it has been shown that acetic acid can enter the yeast cells and lead to intracellular acidification, anion accumulations and inhibition of cellular metabolic pathways (Casal *et al.*, 1996). Acetic acid can also influence cell viability and lead to programmed cell death (PCD), a process which is similar to mammalian apoptosis (Silva *et al.*, 2013).

It has been previously shown that deletion of *HSC82* and *HSP82* alter sensitivity to acetic acid (Silva *et al.*, 2013). In order to identify novel genes potentially involved in response to stress condition induced by acetic acid, we have identified 7 candidates mutants (*ATP22* (translation activator), *MRPS16* (mitochondrial ribosomal protein), *PCS60* (oxalate catabolic process and mRNA binding), *RPL36A* (ribosomal protein), *VPS70* (unknown), *YBR063CA* (unknown) and *DOM34* (ribosomal subunits disassociation)) which show sensitivity to acetic acid treatments.

Presented in Figure 3-1, spot test (serial dilution) analyses of selected mutants indicated their sensitivity to acetic acid compared to the *WT* (Wild Type). In order to confirm that observed sensitivities are not the side effect of secondary mutations due to the gene deletion, deleted genes were reintroduced into corresponding mutants. Spot test analysis under acetic acid condition showed the sensitivity at *WT* level for reintroduced mutants confirming that the deleted target genes are responsible for the original observed sensitivities.

Next, we investigated the potential role of our candidate genes in Hsp82 and Hsc82 gene expression. In order to investigate whether or not this effect is at the transcriptional level, we performed qRT-PCR analysis for endogenous Hsp82 and Hsc82 expressed within our selected candidates. Presented in Figure 3-2, qRT-PCR analysis showed that deletions of our selected candidates have no significant effect (statistically not significant) on HSP82 and HSC82 at the transcriptional level when compared to the *WT* strains.

Figure 3-1: Spot test analysis for selected candidates under acetic acid condition (220 mM for 2 hours of treatment). Reintroduction of the deleted gene converted the sensitive phenotype to the *WT* sensitivity level. p416 plasmid is used as an empty negative control. All drug sensitivities are performed in triplicate with similar results.

Figure 3-2: *HSP82* and *HSC82* mRNA content analysis. *HSP82* and *HSC82* mRNA contents are related to those of control strains. *PGK1* mRNA content was used for normalization. There are no statistically significant (P-value ≤ 0.05) differences in mRNA contents between *WT* and tested mutants.

To further investigate whether or not our selected candidate genes have an impact on Hsp82 and Hsc82 expression at the translational level, we used 3 plasmid based reporter systems (LacZ expression analysis); p281-4-HSC82-LacZ, p281-4-HSP82-LacZ (Silva *et al.*, 2013) which carry HSP82 and HSC82 regions fused to the LacZ expression cassette, respectively, and p281-4-CLN3-LacZ (Silva *et al.*, 2013) as a negative control. Of interest, presented in Figure 3-3A, our mutants showed lower HSP82/HSC82-mediated β -galactosidase activity when compared to the *WT* strain. It is possible for alteration in mRNA levels to explain differences in gene expression observed in this plasmid based system. To examine this possibility we investigated the mRNA content via qRT-PCR. No statistically significant differences between the mRNA content of these mutants compared to the *WT* was observed (Figure 3-3B). This indicates that the observed decrease in the expression of HSP82/HSC82-mediated β -galactosidase appears to be at the protein synthesis level and not due to a decreased in mRNA content. In agreement with these observations, it has been previously shown that induction of apoptosis by acetic acid in yeast has a direct impact on the IRES-mediated translation. Both HSP82 and HSC82 suggested having IRES elements within their sequences (Almeida *et al.*, 2009, Silva *et al.*, 2013).

To further investigate whether or not our selected candidate genes impact general IRES-mediated translation, we used another plasmid, p281-4-URE2-LacZ (Komar *et al.*, 2003), that carries an IRES element that belongs to *URE2* gene fused to LacZ expression cassette. Presented in Figure 3-3A, the majority of our selected mutants showed a lower level of IRES-mediated β -galactosidase activity when compared to the *WT* strain. qRT-PCR analyses indicated that observed alteration in LacZ expression cassette appears to be

at the translational level and not at the level of mRNA content. No significant difference was observed between *WT* and mutants for the expression of the negative control plasmid (p281-4-CLN3-LacZ, contains no IRES sequences fused into LacZ). Note that all of these constructs have the same plasmid backbone in which cap-dependent translation is blocked through four inhibitory hairpin structures (Komar *et al.*, 2003; Silva *et al.*, 2013).

Figure 3-3: A) β -galactosidase expression from templates that contain an IRES element fused to the LacZ expression cassette. β -galactosidase expression level of the mutants are normalized to the expression level of *WT* set at 1. *(P-value ≤ 0.05) and ** (P-value ≤ 0.01). B) β -galactosidase mRNA content analysis. β -galactosidase mRNA content are related to those of control strains. *PGK1* mRNA content was used for normalization. There are no statistically significant differences in mRNA contents between *WT* and tested mutants. The averages are obtained from at least 3 independent experiments.

Altogether these observations provide evidence that our target genes appear to affect IRES-mediated protein synthesis. Of interest, most of them influence the expression of all three tested IRES elements to certain degree. However, the levels at which each IRES affect translation is very different. For example, deletion of *ATP22* demolishes the expression mediated by Hsc82 IRES. However it reduces the expression mediated by Hsp82 and URE2 IRESs by only ~15 and 50%, respectively. Deletion of none of the studied genes abolished translation of all three IRESs.

3.4.2 Genetic interaction (SGA and PSA) investigations

To better investigate and provide further evidence for the involvement of *ATP22*, *MRPS16*, *PCS60*, *RPL36A*, *VPS70*, *YBR063C* and *DOM34* in the translation pathway, we studied the genetic interactions they made with genes that influence the protein synthesis pathway. Synthetic genetic interaction analysis (SGA) is a high-throughput genetic interaction analysis technique based on mating a reference mutant strain (mating type α) to each mutant in the yeast haploid gene deletion array (mating type a) followed by haploid double mutation selection. Selected double mutants are then scored for genetic interaction using growth defect (Tong *et al.*, 2001; Memarian *et al.*, 2007; Jessulat *et al.*, 2008; Samanfar *et al.*, 2013; 2014). Genetic interaction can be explained by the fact that alteration in expression of two genes (double mutant) results in a phenotype which cannot be justified by the alteration of individual gene expression. Such genetic interactions are commonly used to investigate the functional relation of two genes. The most commonly studied form of genetic interaction is a negative genetic interaction where double mutants have a lower growth rate (sick or lethal) than the expected individual mutant growth phenotype. These interactions often disclose genes that are functionally related through

parallel pathways (overlapping pathways). In parallel pathways, one gene/pathway can compensate the activity of the other. These interactions may provide insight into a higher level of association between gene functions. In this way, function of uncharacterized gene(s) may be investigated by the genetic interactions they made with other genes with known functions (Tong *et al.*, 2001; Samanfar *et al.*, 2013; Omid *et al.*, 2014; Samanfar *et al.*, 2014). To this end, we investigated the sick interactions that *atp22Δ*, *mrps16Δ*, *pcs60Δ*, *rpl36aΔ*, *vps70Δ*, *ybr063cΔ* and *dom34Δ* formed with two sets of 384 gene deletion strains. Set one, called translation array, contains known genes involved in translation array and set two, called random array, carries random gene deletions (excluding those involved in translation pathway) used as a control. In this way, 768 double mutants are systematically generated for each gene in triplicate (16128 double deletions in total). Some genes may only form interactions under specific conditions, for example during DNA damage (Omid *et al.*, 2014). Consequently, their interactions would not be observed under standard laboratory growth condition. To have a better coverage for such conditional interaction, we repeated our genetic interaction analysis under targeted stress conditions including heat shock, acetic acid treatment and the presence of translation inhibitory reagents. In all cases, the growth fitness of double mutant gene deletions was quantified by colony size measurements (Memarian *et al.*, 2007; Samanfar *et al.*, 2013; 2014) and color-coded. To improve coverage, we combined our interaction data with those previously reported (<http://drygin.ccbr.utoronto.ca>) (Figure 3-4). A complete list of interactions is presented in Appendix 3-2. To have better understanding of what these genetic interactions may imply, we performed gene ontology (GO) annotation analysis on the genetic interacting partners of our target genes. In this

way we would evaluate the enrichment of cellular function/process for the interacting partners (P-value: ≤ 0.05). Note that reintroduction of *ATP22*, *MRPS16*, *PCS60*, *RPL36A*, *VPS70*, *YBR063C* and *DOM34* into a representative set of corresponding double mutants, driven from SGA, reversed the sick phenotype observed for the double mutants, further confirming that the observed sick phenotypes are caused by the deletion of the target genes of interest and not by a possible secondary mutation within the genome (Figure 3-5 and Appendix 3-3).

Figure 3-4: The representative SGA analysis for *VPS70*; the combination (from this study and the literature) of synthetic sick interaction for *VPS70*. *VPS70* interacts with genes with the following gene ontology: structural constituent of ribosome (P-value: 5.30E-18) and regulation of translation (P-value: 3.71E-12). *Represents interactions that were included from literature. ^A Represents conditional SGA under acetic acid treatment (180 mM). ^C Represents conditional SGA under cycloheximide (40 ng/ml) treatment. ^H Represents conditional SGA under heat shock (37 °C) condition. ^P Represents conditional SGA under paromomycin (18 mg/ml) condition. ^R Represents conditional SGA under rapamycin condition (4 ng/ml). ^S Represents SGA data under standard laboratory condition.

Structural constituent of ribosome
(P-value: 5.30E-18)

Regulation of translation
(P-value: 3.71E-12)

20% 95%
Degree of sickness (colony size reduction)

Other

Figure 3-5: Re-introduction of deleted target genes in double mutants. Re-introduction of target genes reverses the observed genetic interactions phenotype. The sick phenotypes of double gene knockout strains are reversed when target genes are re-introduced into the corresponding mutant strains.

Next we investigated the ability of the overexpression of *ATP22*, *MRPS16*, *PCS60*, *RPL36A*, *VPS70*, *YBR063C* and *DOM34* genes to compensate the sick phenotype of different gene deletion strains in response to heat shock, acetic acid, cycloheximide, paromomycin and rapamycin treatments (PSA, phenotypic suppression analysis). Note that PSA is a similar approach to the SGA and explores genetic interactions with the exception that the compensatory effect of the overexpressed target gene is being investigated. If the overexpression of a target gene can compensate the phenotype caused by the absence of another partner gene, a functional connection between the two genes will be considered (Sopko *et al.*, 2006, Alamgir *et al.*, 2008; Omid *et al.*, 2014; Samanfar *et al.*, 2014). To this end (similar to SGA), the single gene deletion haploid strains in the two gene deletion array described above (translation array and random gene array) were systematically and separately transformed with overexpression plasmids *ATP22*, *MRPS16*, *PCS60*, *RPL36A*, *VPS70*, *YBR063C* and *DOM34* in addition to an empty plasmid (p416) as negative control. Single gene deletion strains along with those carrying overexpression plasmid were grown in the presence of a sub-inhibitory concentration of acetic acid (220mM), cycloheximide (60ng/ml), paromomycin (22mg/ml), rapamycin (6ng/ml) and heat shock (37 °C). Those gene deletions which show sensitivity to the treatments, the sensitivity of which were compensated in the presence of an overexpression plasmid were selected as positive hits on the basis of colony size measurement (Memorial *et al.*, 2007; Samanfar *et al.*, 2014; Omid *et al.*, 2014). Positive hits were reconfirmed via spot test analysis (Figure 3-6).

Figure 3-6: Representative spot test confirmation for phenotypic suppression analysis of *VPS70*. Gene deletion mutants for *RPL34A* and *TEF4* show sensitivity to acetic acid (220mM). Overexpression of *VPS70* compensates for the observed sensitivity.

The list of mutants which showed compensation in the presence of overexpression plasmids under stress conditions are presented in Appendix 3-4. Of interest, we observed significant enrichment of genetic interactions for all of our 7 target genes, mainly with genes involved with ribosome biogenesis and regulation of translation. The P-values for these enrichments ranged from 1.28E-19 to 1.14E-05. These observations further connect the activity of our target genes to the process of translation and confirm their involvement in higher levels of communication within the process of protein synthesis.

ATP22 is a specific translational activator for the mitochondrial ATP6 mRNA, deletion of which has sensitivity to acetic acid. Atp22 is translated in the cytoplasm and then exported into the mitochondria. SGA analysis has shown its genetic interaction with genes involved in regulation of translation (P-value: 2.08E-11) and ribosome biogenesis (P-value: 1.28E-19). Its overexpression has compensated the sick phenotype of genes involved in ribosome biogenesis (P-value: 2.74E-15) and regulation of translation (P-value: 2.23E-12) as well. Interestingly, *ATP22* interacts with *NUP2* and *NUP84* that code for proteins involved in nucleopore complex which are implicated in the export of mRNAs from nucleus in response to the stresses including heat-shock. *β-galactosidase* analysis has shown that *atp22Δ* has lower IRES-mediated translation of LacZ expression cassette in URE2 and HSC82. Considering the fact that mitochondrion is directly tied to stress response, it is not unexpected to observe an involvement for *ATP22* in stress-induced IRES-mediated translation.

Mrps16 is a mitochondrial ribosomal protein of the small subunit. *mrps16Δ* showed sensitivity to acetic acid and has lower *β-galactosidase* activity for all three of the IRES-

mediated LacZ expression systems used in this study. Genetic interaction analysis showed that it has negative genetic interactions with genes involved in the regulation of translation (P-value: 1.14E-05) and structural constituents of ribosome (P-value: 2.59E-06); its overexpression compensated the absence of genes which have the same gene ontology categories as SGA analysis, further connecting its activity to cytoplasmic translation. *MRPS16* was found to have negative genetic interactions with *ANB1* (gene encoding an eIF5A protein), having a function in both initiation and elongation steps of protein synthesis. In light of the data in the current study, it is possible that this protein plays a dual role in both mitochondrial translation (as suggested by literature) and cytoplasmic translation.

PCS60 is a gene which produces peroxisomal protein that binds mRNA; SGA data indicate that it interacts with number of genes involved in rRNA metabolism and ribosome biogenesis (P-value: 2.65E-09). Analyzing these interactions along with the co-localization and co-expression data suggest that this protein might play an intermediate role binding ribosomal proteins to the RNA. In agreement with this putative role, PSA analysis has shown that it compensates for the absence of genes involved in structural constituents of the ribosome (P-value: 9.09E-04). *β-galactosidase* analysis has indicated that *pcs60Δ* has lower LacZ enzyme activity mainly for URE2 construct. Since it has a RNA binding domain and its protein product colocalizes with proteins involved in response to the stress, we may conclude its functional role in IRES-mediated translation under stress condition and that this activity appears to be IRES specific and not general.

RPL36A is a gene that codes for the 60S ribosomal subunit protein, L36A, which has an RNA binding domain. It is an important gene needed for proper functioning of the ribosome. Due to its RNA binding ability and reduction of IRES-mediated translation when deleted, a probable role for this protein might be a physical binding to IRES elements to perform IRES-mediated translation initiation. SGA analysis has indicated its interactions with numerous genes involved in ribosome biogenesis (P-value: 7.45E-11). PSA data are in good accord with SGA analysis connecting the activity of L36A to ribosome biogenesis and hence translation.

VPS70 is a gene of unknown function thought to be involved in vacuolar protein sorting. It has negative genetic interactions with genes involved in ribosome biogenesis (P-value: 5.30E-18) including *LTV1*, *RPS16A* and *RPS17A*. In accord with SGA analysis, PSA investigation also showed that overexpression of *VPS70* compensated for the absence of genes involved in ribosomal biogenesis (P-value: 9.60E-14) and regulation of translation (P-value: 1.24E-11). Interestingly, like *ATP22*, it has negative genetic interaction with *NUP84*, nucleopore complex which is implicated in the export of mRNAs from nucleus in response to the stresses

YBR063C is a gene with unknown function, deletion of which causes sensitivity to acetic acid. *ybr063cΔ* has significantly lower IRES-mediated β -galactosidase activity for URE2 and HSP82 IRESs. *YBR063C* negatively interacts with genes involved in biogenesis of ribosome including *RPL43A*, *RPS10A*, *NOP16* and *RPS9B*, as well as *RPL43A*, *RPS10A* and *RPS9B* (P-value: 1.76E-10). In light of the observed data we propose a name change for *YBR063C* into *IAE1*, IRES Associated Element 1.

Finally, *Dom34p* is a protein which facilitates ribosomal subunit dissociation, has endoribonuclease activity and RNA binding domain and is sensitive to acetic acid when deleted. *dom34Δ* has shown low IRES-mediated β -galactosidase activity for all the three constructs indicating its direct impact on IRES-mediated gene expression (to a lesser impact on URE2 IRES). It has a negative genetic interaction with *GSY1*, glycogen synthase, which has been implicated in response to various types of stresses further connecting its observed effect on translation to stress response.

4 Chapter: Efficient prediction of human protein-protein interactions at a global scale

4.1 Abstract

Our knowledge of global protein-protein interaction (PPI) networks in complex organisms such as humans is hindered by technical limitations of current methods. On the basis of short co-occurring polypeptide regions, we developed a tool called MP-PIPE capable of predicting a global human PPI network within three months. We predicted 172,132 putative PPIs and demonstrate the usefulness of these predictions through a range of experiments, especially identification of new players involved in translation pathway. The speed and accuracy associated with MP-PIPE can make this a potential prediction tool to study individual human PPI networks (from genomic sequences alone) for personalized medicine.

4.2 Introduction

Protein-protein interactions (PPIs) are essential molecular interactions that define the biology of a cell, its development and responses to various stimuli. Physical interactions between proteins can form the basis for protein functions, communications, and regulation and controls within a cell. Such interactions can result in the formation of protein complexes that perform specific tasks. Similarly, internal and external signals are often realized and communicated through the formation of stable or transient PPIs. Due to their central importance to the integrity of communication networks within a cell, PPIs

are thought to involve important targets for drug discovery (Khan *et al.*, 2011) and are linked to a number of cellular conditions and diseases (Nibbe *et al.*, 2011).

Our current knowledge of global PPI networks in different organisms is hindered by the constraints and limitations of existing experimental techniques amenable to high-throughput PPI studies, such as yeast-two-hybrid (Y2H) and affinity purification combined with mass spectrometry (APMS). While both of these techniques have been successfully applied to global PPI detection in the yeast, *Saccharomyces cerevisiae* (Uetz *et al.*, 2000; Ito *et al.*, 2001; Gavin *et al.*, 2006; Krogan *et al.*, 2006), they suffer from significant shortcomings highlighted by the lack of overlap observed between the PPI data in different reports. The two benchmark large-scale yeast APMS investigations have less than 25% overlap and this overlap is even less for the two classic Y2H projects (Jessulat *et al.*, 2011). Only 24 PPIs are shared between all four studies, further highlighting the gap in our understanding of global PPI networks. Although recent technical improvements are expected to increase the confidence of the detected PPIs and hence fill some of the current gap of knowledge, increasing the coverage and quality of PPI networks remains an important challenge (Ito *et al.*, 2001; Von Mering *et al.*, 2002; Qi *et al.*, 2006; Lievens *et al.*, 2009; Jessulat *et al.*, 2011).

Computational tools offer time and cost effective alternatives to traditional wet-lab PPI detection tools. They may also be used as “filters” to increase confidence in data derived from wet-lab experiments (Pitre *et al.*, 2008a; Jessulat *et al.*, 2011). Like other techniques, most computational tools also suffer from notable deficiencies. For example, most computational methods rely heavily on previously reported data. Assuming that

there are inherent discrepancies in the training data, the accuracies of such tools to detect new interactions are often questionable. Moreover, novel interaction domains or motifs are likely to be missed by methods that rely heavily on the structures or other high-level features of protein pairs known to interact. Another major shortcoming of computational tools is that they are often too computationally intensive, making them impossible to use for proteome-wide analysis. To date, no comprehensive all-against-all analysis of the entire human PPI network has been possible.

A small number of large-scale computational PPI prediction methods have recently been published (McDowall *et al.*, 2009; Elefsinioti *et al.*, 2011; Zhang *et al.*, 2012). Although these methods have provided important contributions to the field, they are not applicable to the entire human proteome due to computational complexity, availability of input protein features, or unacceptably high false positive rates. For example, a recent study by Elefsinioti *et al.*, 2011, examined five million protein pairs and predicted 94,009 “high confidence” interactions. Given a conservative estimate of 22,000 human proteins, leading to 242 million possible pairs; they have examined only 2% of the potential interactome while others have examined just over 7% (McDowall *et al.*, 2009) and 12.4% (Zhang *et al.*, 2012) of the total interactome. Presumably these methods were limited to examining only small subsets of protein pairs due to computational complexity (i.e. runtime) or the availability of input protein features. For example, the method developed by Elefsinioti *et al.*, 2011, requires 18 complex features for each protein relating to annotated function, sequence-derived attributes, and network structure. Likewise, the method of Zhang *et al.*, 2012 requires structural information for both proteins in the putative interaction and is therefore only applicable to 13,000 human proteins. When

considering protein pairs rather than individual proteins, approximately 50% sequence coverage results in an examination of at most 25% of the possible PPIs. In fact, Zhang *et al.*, 2012, report that they were able to develop models for 36 million interactions, representing 12.4% of the 242 million possible interactions. Even if these methods could be applied to all human protein pairs, typical false positive rates will render existing methods unusable on larger data sets. For example, considering that the method of Elefsinioti *et al.*, 2011 predicts 94,009 “high confidence” interactions among only 1.6% of protein pairs, then we can reasonably expect nearly 6 million “high confidence” predicted interactions if their method were to be applied to the entire human proteome. This is an order of magnitude higher than the largest current estimate of the true size of the human interactome, leaving the experimenter to weed through a multitude of false positive predictions to find the few true interactions. Likewise, using a previously published computational method (Zhang *et al.*, 2010a), Zhang *et al.*, 2012 recently reported a false positive rate implying 41.2% precision, and their recall over an independent test set of 24,000 newly reported PPIs is less than 7%. Consequently, there is a need for the development of efficient tools that are readily amenable to proteome-scale PPI prediction. This is especially important as the field of personalized medicine will benefit tremendously from a fast and accurate method that can predict the global PPI maps of different individuals from their genomic sequences alone.

A subset of cellular PPIs is mediated by defined short, linear polypeptide sequences (Neduva and Russell, 2006; Chica *et al.*, 2009; Stein and Aloy, 2010). Leveraging this fact, a number of computational tools have been developed to detect PPIs solely on the basis of primary sequence (Pitre *et al.*, 2006; 2009; Petsalaki *et al.*, 2009). Such

approaches do not rely on known structures or other protein features that are not easily deduced from primary protein sequences, and are thus, in principle, able to interrogate portions of the proteome that are inaccessible to other methods. Some of their predictions have been confirmed by tandem affinity purification (Pitre *et al.*, 2006), *in vitro* binding assays (Neduva *et al.*, 2005), and *in vivo* functional analysis (Pitre *et al.*, 2008a; 2008b). An added benefit of sequence-based PPI prediction is that short polypeptide sequences in one organism can be used to predict PPIs in another (Pitre *et al.*, 2012). We note that, while the wide applicability of sequence-based PPI prediction methods is clearly strength, in not using structural predictions, such techniques may be unable to account for structural features such as binding site accessibility or widespread contacts between non-contiguous residues.

We have developed a computational tool termed the Protein Interaction Prediction Engine (PIPE) that uses co-occurrence of short polypeptide regions to detect novel PPIs in *S. cerevisiae* (Pitre *et al.*, 2006). Although PIPE was able to analyze potential PPIs within certain proteomes, applying this tool to more complex proteomes remained infeasible due to computational complexity. Analyzing the ~242 million protein pairs in the human proteome was estimated to require approximately 6.3 million CPU-hours of computation. In order to study the human PPI network, we developed a new Massively Parallel (MP) version of PIPE, which we call MP-PIPE. MP-PIPE overcomes some of the limitations of existing methods through computational acceleration of the algorithm (speed) and improved precision. We present a comprehensive all-against-all (pair-wise) analysis of the human proteome and study its biological properties. We then demonstrate

the accuracy and utility of the MP-PIPE inferred interactome using a range of functional assays such as translation pathway related investigations.

4.3 Materials and methods

4.3.1 Biological experiments

4.3.1.1 Yeast growth condition

Standard rich (YPD) and synthetic complete (SC) media were used as a growth media (Alamgir *et al.*, 2008). To investigate translation inhibitory drugs (antibiotics) on growth rate of yeast deletion mutants, streptomycin (40 mg/ml) was added to SC media and cycloheximide (60 ng/ml) was added to YPD media (Alamgir *et al.*, 2008).

4.3.1.2 Drug sensitivity test

Yeast cells were grown in liquid YPD and SC media to mid-log phase and diluted to a concentration of 10^{-2} to 10^{-5} cells/20 μ l. From each dilution, 20 μ l was spotted onto solid media containing translation inhibitory drugs. Media with no antibiotics was used as a control. All yeast cells were incubated at 30°C for 1-2 days (Jessulat *et al.*, 2008).

4.3.1.3 Protein expression assay

Translation fidelity was measured using plasmids pUKC817, pUKC818 and pUKC819, which carry premature stop codons UAA, UGA and UAG within a β -galactosidase expression cassette (Lucchini *et al.*, 1984; Alamgir *et al.*, 2008). Translation efficiency was assessed using plasmid p416 (containing Gal-inducible LacZ expression cassette) as described by (Krogan *et al.*, 2003; Alamgir *et al.*, 2008). β -galactosidase assay was

performed using ONPG, O-nitrophenyl- α -D-galactopyranoside, as described by (Stansfiels *et al.*, 1995; Krogan *et al.*, 2003; Shenton *et al.*, 2006) in three repeats.

4.3.1.4 qRT-PCR

Total RNA was extracted using RNeasy Mini Kit (QIAGEN) according to manufacturer's instruction. To synthesize cDNA, 33 μ l of RNA in combination with 3 μ l poly T primer (random hexamer) was incubated for five min at 70°C, and then cooled on ice for five min. 6 μ l RT-buffer, 15 μ l dNTPs, and 3 μ l RNaseI were added to the mixture and incubated for five min at 37°C. 3 μ l RT enzyme was added to the mixture and incubated for an additional 1hour at 42°C followed by 10 min incubation at 70°C. qPCR amplification and detection was performed on a Rotor Gene 3000 from Corbett Research. A final mixture of 2.4 μ l H₂O, 2.4 μ l 10X PCR buffer, 1,25 μ l dNTPs (4mM), 1.25 μ l MgCl₂, 2 μ l SYBR Green (1/4000), 7.6 μ l primer mix (1mM) and 3 μ l Taq was used in addition to 5 μ l template (cDNA). Thermo-cycler conditions were set to the following: 50°C for two min, 95°C for 10min, 40 cycles of 95°C for 30s-60°C for 30s-72°C for 30s and a final 72°C for ten min (Pfaffl *et al.*, 2001; Yu *et al.*, 2007b). The primers used in the qRT-PCR were designed based on the sequences for LacZ (F: TTGAAAATGGTCTGCTGCTG, R: TATTGGCTTCATCCACCACA) and PGK-1 (F: CAGACCATTCTTGGCCATT, R: CGAAGATGGAGTCACCGATT). PGK-1 (phosphoglycerate kinase) was selected as the positive control due to being one of the most highly expressed genes in yeast, producing up to 5% of total mRNA content (Chambers *et al.*, 1989).

4.3.1.5 Versatile affinity-tagging, purification, and protein identification

The full length, sequence verified, non-mutated, Gateway-compatible cDNA entry clones for the human chromatin-related proteins (CBX1, RNF2, H2AFX, and RBBP4) were obtained from Harvard Plasmid ID. The HEK293 and HEK293T cells were cultured in Dulbecco's modified Eagle's medium with 10% fetal bovine serum and antibiotics essentially as previously described (Mak *et al.*, 2010; Ni *et al.*, 2011). The sequence-verified clones were cloned into the lentiviral expression vector essentially as previously described by (Mak *et al.*, 2010; Ni *et al.*, 2011). The lentivirus-encoded tagged ORFs were transduced into HEK293 cells, and the stably expressed cells were subsequently selected with puromycin at a concentration of 2 µg/ml for a minimum of 48 hours and expanded, essentially as described previously (Mak *et al.*, 2010; Ni *et al.*, 2011). The expression of the tag in stable cells was subsequently confirmed by Western blotting using anti-FLAG antibody against the 3X FLAG epitope. Affinity purification and sample processing for protein identification by mass spectrometry was performed essentially as previously described by (Mak *et al.*, 2010; Ni *et al.*, 2011). The high-confidence matches of the resulting MS/MS spectra was mapped to the reference human protein sequences using the SEQUEST database search engine with match quality evaluated using the STATQUEST algorithm (Kislinger *et al.*, 2006). The identified co-purifying proteins were filtered out at a confidence threshold at 90% with two or more peptides. Since each tagged samples were independently affinity purified two times to rule out the non-specific binding proteins, we averaged the peptide counts over the replicates. Moreover, the co-purifying proteins that were identified at 90% cut-off with one single peptide and the most common background contaminants or proteins that bound

to the unrelated VA-tagged GFP samples in our replicate purifications were filtered out to eliminate the noise from the dataset.

The comparison of LGTS-MS results with the MP-PIPE predicted and previously reported (known) interactions was done by calculating the precision and recall on the basis of LGTS-MS results representing real interactions. To do this we filtered the MP-PIPE predictions and known interactions to contain detectable proteins (the set of all proteins seen in any of the LGTS-MS experiments performed here). Reachable proteins were defined as those proteins that interact directly with the bait or indirectly through one or two intermediaries within the sets of filtered interactions.

4.3.1.6 Prediction and analysis of candidate biomarkers for seasonal allergic rhinitis, SAR

40 patients with SAR were included in the study. SAR and symptom scores were defined as previously described (Benson *et al.*, 2000; Wang *et al.*, 2011a; 2011b). Their median (range) age was 28 (18-58) and 19 were women. The mean $\pm$ SEM symptom score of the 40 patients after treatment decreased from 15.7 ± 1.0 to 4.3 ± 0.6 ($P \leq 0.0000001$). The study was approved by the Ethics Committee of the Medical Faculty of the University of Gothenburg. Written informed consent and questionnaire data sheets were obtained from all patients. Proteins from the acute phase response pathway, complement signaling pathway, and glucocorticoids receptor pathway were extracted from the Ingenuity pathway analysis (IPA) software. Interactors to these proteins were predicted using PIPE. We included secreted, membrane and cytoplasmic proteins, but excluded nuclear proteins. We prioritized candidate biomarkers based on their number of known and

predicted interactions. Next, we focused on proteins with a high number of predicted interactions.

Proteins were examined by ELISA in nasal fluids from 40 patients with SAR before and after GC treatment. V-akt murine thymoma viral oncogene homolog 2 (Akt2) was examined with an ELISA kit from R&D Systems Inc. (Minneapolis, MN, USA). Proline-rich protein BstNI subfamily 1 (PRB1), proline-rich protein BstNI subfamily 2 (PRB2), v-yes-1 Yamaguchi sarcoma viral related oncogene homolog (LYN) and stratifin (SFN) were analyzed with ELISA kits from Usclife Life Sciences and Technology (Wuhan, China). All experiments were performed according to the manufacturer's protocol. The Wilcoxon matched pairs signed ranks test was performed to compare two paired groups. A P-value less than 0.05 was considered significant.

4.3.2 Computational

4.3.2.1 Sequential PIPE algorithm

For a given organism (e.g. *S. cerevisiae*, *C. elegans*, or human (*Homo sapiens*)), the PIPE algorithm relies on a database of known and experimentally verified protein interactions. For example, for the 22,513 human potential open reading frames included in the current study, only 41,678 high confidence interactions are known (out of 253,406,328 possible protein pairs). Since experimental verification can have large numbers of false positives (up to ~40%, see Pitre *et al.*, 2006), the PIPE database is carefully constructed to avoid false data and stores only protein interactions that have been independently verified by multiple experiments. The database represents an interaction graph G where every protein corresponds to a vertex in G and every interaction between two proteins X and Y is

represented as an edge between X and Y in G . The remainder of this section outlines how, for a given pair (A, B) of query proteins, our PIPE method predicts whether or not A and B interact. In the first step of the PIPE algorithm, protein A is split up into overlapping fragments of size w . This can be thought of using a sliding window of size w across protein A . For each fragment a_i of A , where $0 \leq i \leq |A| - w + 1$, we search for fragments "similar" to a_i in every protein in graph G . A sliding window of size w is again used on each protein in G , and each of the resulting protein fragments is compared to a_i . For each protein that contains a fragment similar to a_i , all of that protein's neighbors in G are added to a list R . To determine whether two protein fragments are similar, a score is generated with the use of the PAM120 substitution matrix. If the similarity score is above a tunable threshold then the fragments are said to be similar. In the next step of the PIPE algorithm, protein B is split into overlapping fragments b_j of size w ($0 \leq j \leq |B| - w + 1$) and these fragments are compared to all (size w) fragments of all proteins in the list R produced in the previous step. We then create a result matrix of size $n \times m$, where $n = |A|$ and $m = |B|$ and initialize it to contain zeroes. For a given fragment a_i of A , every time a protein fragment b_j of B is similar to a fragment of a protein Y in R , the cell value at position (i, j) in the result matrix is incremented. The result matrix indicates how many times a pair (a_i, b_j) of fragments co-occurs in protein pairs that are known to interact. It is based on this matrix that the query proteins are predicted to interact or not. A modified median filter, which simply sets a cell's value to 1 if most of the neighboring cells are greater than zero and zero otherwise, is applied and the two query proteins were predicted to interact if the average cell value was above a set threshold. By varying this threshold, a range of precision-recall values may be obtained (Appendix 4-1). Note that throughout

this paper, for our analysis a prevalence of 1 PPI per 100 protein pairs is consistently assumed for our results, as well as for comparison to other results as was done in (Park, 2009). Recall measures the proportion of true interactions that will be detected. Precision measures the proportion of predicted interactions that correspond to true interactions. For our LOO experiments, our 41,678 high confidence positive PPIs are taken from BioGrid (Stark *et al.*, 2006). Random protein pairs not previously reported to interact were used for our negative interaction data. This is considered to be a conservative approach when assessing prediction accuracy (Yu *et al.*, 2010).

4.3.2.2 MP-PIPE overview

The MP-PIPE (massively parallel PIPE) system is a massively parallel, high-throughput protein-protein interaction prediction engine and is the first system that is capable of scanning the entire protein interaction network of complex organisms such as human. In order to achieve that goal, large numbers of concurrent PIPE instances need to be executed on a large-scale parallel compute cluster. This created two major challenges. The first problem was the lack of scalability that made it difficult for large numbers of PIPE instances to effectively take advantage of all available computational resources without massive load imbalances. This load-balancing problem was not as significant in simpler organisms, such as *S. cerevisiae* and *C. elegans*, but lead to a large amount of unused resources when making predictions on more complex organisms such as human. In particular, the calculation/prediction of these interactions is considerably more time consuming. Previous PIPE experiments for *S. cerevisiae* (Pitre *et al.*, 2006; 2008a) and experiments for *C. elegans* reported in (Pitre *et al.*, 2012) showed that PIPE can process

each individual protein pair within seconds. However, for human proteins, the picture changes dramatically. The running time for one individual protein pair can fluctuate between less than a second and more than 12 hours. Human proteins have a much more complex structure that appears to lead, in some cases, to a very large number of fragment similarities found by PIPE. When trying to run earlier versions of PIPE on human protein pairs, individual PIPE instances would simply be given static lists of protein pairs to make predictions on. Due to the wide variance of processing time for human protein pairs, some PIPE instances would finish very quickly while, by the end, there may be a single PIPE instance working for hours on a single protein pair while all of the other instances are idle. The imbalance when processing human protein pairs was so great that it resulted in more wasted resources than utilized resources when processing batches of protein pairs. To process a global scan of all human protein pairs, this issue had to be overcome.

The second major issue facing concurrent PIPE instances on a processor is inefficient usage of memory. Typically the number of PIPE processes running on a single machine is set to the number of compute cores on that machine. For example, on a quad-core machine there would typically be four PIPE processes running to utilize the chip fully. If different PIPE processes were left to work completely independently of each other, each process would have to load its own copy of the interaction graph along with all the other PIPE data. For less complex organisms this was not a major issue since the amount of data loaded was relatively small but the complexity of the human proteome translates into significantly more data needed by PIPE. The memory needs for a single PIPE instance for the human proteome increased to such a degree that running as many PIPE instances

as compute cores can easily lead to program crashes due to a lack of memory. This would imply that processor cores would be left unused due to memory limitations. To process a global scan of all human protein pairs, this issue had to be overcome. The basic structure of MP-PIPE is a two-level master/slave and all-slaves model. A single MP-PIPE scheduler process is in charge of managing the main list of protein pairs to be processed as well as reporting the results. The MP-PIPE scheduler distributes work to several MP-PIPE worker processes in packets. Each packet contains a relatively small number of protein pairs. Each MP-PIPE worker executes the PIPE algorithm on protein pairs received from the MP-PIPE scheduler. By giving each worker only a relatively small amount of work at a time we ensure that if a worker does get stuck with abnormally time consuming protein pairs, the other workers will continue to work on their packets and, when they finish, they will request more work from the scheduler process and continue to work. This aspect of the MP-PIPE's architecture deals with the load imbalance problem by ensuring that all PIPE processes are working as long as there is still work to be done. It should be noted however that if the packet size is too small then the amount of communication between the scheduler and worker processes will negatively impact the running time of the system. It is therefore important to balance the packet size between being too small (too much communication overhead) and too large (too much work imbalance). To improve the memory efficiency, the second level of MP-PIPE's architecture uses an "all-slaves" model. Each PIPE worker process consists of a number of parallel threads, called worker threads, among which it distributes the protein pairs to be processed. The worker threads of an MP-PIPE worker are to be executed on a shared memory multi-core processor. The PIPE interaction graph and other necessary PIPE data

require considerable amounts of memory. For MP-PIPE, the data stored at an MP-PIPE worker process was re-designed to become a parallel data structure on which all worker threads for that worker can operate concurrently. Much care was taken to implement this as memory efficiently as possible so that a single shared copy fits into the main memory of a processor node executing an MP-PIPE worker. This allowed more threads to run simultaneously on a given processor node by reducing the overall memory usage and solved the memory issues discussed. The scheduler/worker part of MP-PIPE was implemented using MPI (Message Passing Interface) and the worker threads within each MP-PIPE worker were implemented in OpenMP (<http://openmp.org/>).

4.3.2.3 Complete scan of the human proteome

MP-PIPE evaluated all possible protein pairs in the human proteome. Most of this work was performed on the large *Victoria Falls* cluster, using the medium cluster to offload some of the harder pairs (i.e. those pairs that took longer than 12 hours). The following represents the human proteome at the time that the MP-PIPE scan was performed.

- Total number of human proteins: 22,513
- Total number of protein pairs to examine: 253,406,328
- Total number of known protein interactions: 41,678
- Total number of proteins with at least one known interacting partner: 9,459
- Total number of proteins with no known interacting partners: 13,054
- Largest number of known interactions partners for a single protein: 265
- Average number of known interactions per protein: 3.70
- Average number of known interactions per protein with at least one interaction: 8.81

The human proteome has almost seven times more known interactions than the *C. elegans* proteome. Coupled with the fact that the human proteins are, on average, longer than the *C. elegans* proteins, this significantly increases the complexity of scanning the entire human proteome. Furthermore, as outlined above, the running time for one individual protein pair can fluctuate between less than a second and more than 12 hours. This creates an additional load-balancing problem as discussed above. In fact, some individual protein pairs required six days of computation. On the Victoria Falls cluster, 50 nodes were used, each with their own MP-PIPE worker process running 256 threads. This implies 12,800 parallel computational threads running on 6,400 hardware-supported threads. The number of threads per node was scaled down from 512 threads due to the fact that each individual thread needed significantly more memory than in our tests. The Victoria Falls cluster was used to process the vast majority of protein pairs. The results presented in this paper were obtained through three months of exclusive 24/7 use of the large Victoria Falls cluster. If one of its worker threads got stuck with a protein pair that was running more than 12 hours, that protein pair was off-loaded to the medium cluster since its individual cores are more powerful than a single Victoria Falls thread.

4.3.2.4 Hubs and betweenness centrality

Protein interactions can be represented as an interaction network, where the proteins are interactors (nodes) and connections (interactions) are shown as edges. The number of edges incident to one node is called the degree of that node. Hubs are high degree nodes that interact with many other proteins (nodes) through various pathways (Gursoy *et al.*, 2008). To find the hubs in the human predicted interaction graph, the proteins were sorted

by their degree in the network. Betweenness centrality is an important global property of networks. The betweenness centrality of a node v is the ratio of the number of shortest paths between a pair of nodes a and b on which v lies and the total number of shortest paths between a and b , summed over all possible pairs of nodes. These measures were used to identify proteins of interest. The betweenness centrality of a protein v is defined as $c(v) = \sum_{a \neq v \neq b} \frac{\sigma_{ab}(v)}{\sigma_{ab}}$, where $\sigma_{ab}(v)$ = # shortest paths from a to b through v , σ_{ab} = # shortest paths between a and b .

4.3.2.5 Graph algorithms for assigning protein complexes

To decompose the predicted protein pairs into putative complexes, we applied a novel algorithm that combines pre-existing graph-theoretic tools with hierarchical clustering concepts. The algorithm has three independent stages: the initialization stage, which consists of generating an initial set of clusters, the merge stage, which determines which two clusters to merge next, if any, and the glom stage, which evaluates vertices for inclusion into a cluster. The initialization stage is run once, after which the merge and glom stages run alternately until either the desired number of clusters is reached or until neither stage results in a change to any cluster. Since initialization is an independent step, any initial clustering may be used. It is not required that the initial clustering be overlapping, although stages two and three may grow the clusters so that the end result is overlapping. We chose to use the set of all maximal cliques as the initial clustering. A clique is a set of vertices with all edges present; a clique is maximal if no vertex can be added to it to form a larger clique. The set of maximal cliques forms a natural overlapping clustering of a graph with the most rigid requirements, namely that all edges

be present within each cluster. Real-world graphs often have many small and medium sized maximal cliques, and the protein prediction graph is no exception. These clusters are then allowed to merge and grow in stages two and three, gradually relaxing the stringency until the desired number of clusters is reached. In the merge stage, the overlap of all clusters is evaluated and the two clusters with the highest overlap proportion are merged. If no two clusters overlap by a proportion greater than a parameter m , then no clusters are merged. In the glom stage, every vertex not already belonging to a particular cluster is considered for inclusion into a cluster in similar fashion to the paraclique algorithm described in (Chesler and Langston, 2006). Those vertices with connectivity proportion greater than g , the glom factor, are added to the cluster. The first time through the glom stage, every cluster is considered. Subsequent glom stages only consider the cluster newly created by the merge stage, as all other clusters having already been considered. In practice, calculating all pairwise overlaps to find the highest degree of overlap can make the merge stage computationally prohibitive. A small change, however, yields a good approximation version that can be run until the number of clusters is reduced to the point where the exact version can take over. Rather than merging the clusters with the highest overlap, the approximation version merges the first two clusters encountered with overlap at least a , the approximation parameter. For the protein prediction graph, which was initialized with more than 100,000 maximal cliques, we ran the approximation version until the number of clusters reached 20,000, at which point we switched to the exact version. Ultimately, a list of 8,739 paracliques were identified and characterized through a statistical analysis of the GO annotations of each member protein.

4.4 Results and discussion

4.4.1 MP-PIPE performance and scalability enables computational scan of entire human proteome

One of the main issues when predicting human protein interactions on a large-scale, which does not occur for simpler organisms such as *S. cerevisiae* or *C. elegans*, is the complexity of the human proteome. More precisely, when predicting human protein interactions using previous methods (Pitre *et al.*, 2008a; 2008b), the compute time to process a single human protein pair can vary between several seconds and more than 12 hours. This effect has so far only been observed for the human protein interactions. Our previous method (Pitre *et al.*, 2008a; 2008b) was ranked highly in terms of prediction accuracy in an independent comparison study (Park, 2009). However, it would be unable to process all human protein pairs in our lifetime (approximately 6.3 million CPU-hours of computation). Therefore, we developed an algorithm called MP-PIPE, capable of performing global PPI analysis of the human proteome within three months through massive parallelization. From a Computer Science perspective, the main challenge is the massive load imbalance of the parallelization. As shown in Appendix 4-2-A, for the vast majority of protein pairs, protein interaction prediction can be performed in seconds. However, for some protein pairs, the process takes minutes or hours, more than 12 hours in 8,000 extreme cases. Solving this load imbalance in an efficient manner is the main computational contribution of MP-PIPE. In the following, we discuss the runtime performance of MP-PIPE on different hardware architectures which eventually enabled us to perform global PPI analysis of a human cell within three months. More precisely, we tested our MP-PIPE solution on three different compute clusters. These clusters

included a six node cluster with 24 total compute cores (small cluster), a 32 node cluster with 128 total compute cores (medium cluster), and a 50 node cluster with 6,400 total hardware supported threads (large cluster). The performance of MP-PIPE was initially tested on a single large cluster node with varying numbers of threads, and then in a second test we increased the number of nodes. The test data set consisted of 50,000 random protein pairs. However, this data set proved to be too large to compute using a small number of threads, so a subset containing 5,000 random pairs was used to examine the runtime performance of the code with 1 - 16 threads and then the full 50,000 pair data set was used for tests with 16 or more threads. For those test cases that were performed over the smaller 5,000 pair subset, runtimes were extrapolated to estimate the runtime over the full 50,000 pair dataset. The speedup curve shown in Appendix 4-2-B shows a dramatic performance improvement using up to 128 threads and then a slight improvement from there up to 512 threads. We found that using more than 512 threads creates memory problems. For the second test, we increased the number of large cluster nodes used, where each node ran one MP-PIPE worker with 512 worker threads. The results are shown in Appendix 4-2-C. The performance of MP-PIPE scales almost linearly as the number of compute nodes increases. This scalability property of MP-PIPE enabled us to perform global PPI analysis of the human proteome within three months.

4.4.2 Verification of MP-PIPE against experimental data

To determine the prediction accuracy of MP-PIPE for human PPIs we conducted a leave-one-out test of MP-PIPE against the 41,678 experimentally verified high confidence human PPIs and 100,000 negative protein pairs (assumed to not interact). To estimate the recall of MP-PIPE, we performed 41,678 test runs of MP-PIPE, one for each

experimentally verified interaction for a given protein pair (A, B). In doing so, for each test run (A, B), we removed the known interaction (A, B) from the database. In this manner, we create a state where MP-PIPE is not aware of the experimentally verified PPI (A, B), as if that interaction had not been measured yet. We then asked MP-PIPE to predict whether or not proteins A and B interact. The precision of the method was measured by the false positive rate observed over 100,000 non-interacting protein pairs. By varying the MP-PIPE threshold values which control MP-PIPE's recall (see Materials and Methods for details on these thresholds), a range of precision-recall performance levels can be obtained (see Appendix 4-1 for the full precision-recall curve). At a precision of 82.1% MP-PIPE detected 23% of the 41,678 experimentally verified human PPIs. This leave-one-out test was repeated with all homologs removed from our human dataset as in Park (2009). This reduced our protein sequence set from 22,513 to 14,867 and our experimentally verified human PPI set from 41,678 to 19,588 pairs. As can be seen in Appendix 4-1 (pink line), the recall of our method is slightly reduced when homologous proteins are removed from our dataset, however we still achieve a precision of 69.1% at the same 23% recall.

4.4.3 All-against-all (pair-wise) scan of the human proteome

After three months of 24/7 computation on the 50 fully dedicated nodes of the large cluster (plus additional computation on the medium cluster), MP-PIPE completed the scan of the human proteome. With a recall of 23% and a precision of 82.1%, MP-PIPE predicted 172,132 protein interactions. Of these high confidence predictions, 132,710 protein interactions have never been reported previously. Given that 41,678 human protein interactions are known (previously reported), MP-PIPE has potentially more than

quadrupled our knowledge of the human interaction network. At a recall of 23%, MP-PIPE data covers more than one fifth of the estimated human PPI landscape. In comparison, Elefsinioti *et al.*, 2011 have examined 2% of the interactome while others have examined just over 7% (McDowall *et al.*, 2009) and 12.4% (Zhang *et al.*, 2012) of the total interactome. The list of the reported interactions is found in Appendix 4-3. The list is ordered according to PIPE score, where higher values represent higher confidence levels for an interaction. Distribution of run time for different human protein pairs is illustrated in Appendix 4-2-A.

Besides leave one out cross-validation, another standard method for evaluating PPIs is to check whether the proteins pairs predicted to interact are co-located within the same cellular component, have the same molecular function, are involved in the same biological process or have a common third party interacting partner. The results of this analysis are shown in Table 4-1 and Figure 4-1. The overall profiles for the predicted interacting pairs that have not been detected before, on the basis of cellular localization, process and function resembles that of previously reported pairs (Figure 4-1). Certain differences however are noticeable. For example, a new association for “metal ion binding” and “transcription” is observed for the predicted interactions and not for those that have been previously reported (Figure 4-1B). Similarly, there is a strong association between “immune response” and “signal transduction” for the predicted interactions (Figure 4-2C). As indicated in Table 4-1, the percentage of predicted interacting protein pairs that have similar function, occur in the same cellular component and participate in the same cellular process is 20.6%, which is consistent with the percentage for previously reported protein pairs (35.0%). In contrast, only 0.8% of randomly selected protein pairs

share these three traits. It is important to note that the PIPE algorithm has no previous knowledge of the protein location, molecular function of proteins, or the biological processes in which they are involved. Such an association for protein pairs predicted by MP-PIPE further highlights the ability of this method to predict interactions that can be supported by independent parameters. Such contextual information can also be used to assign an independent degree of confidence for PPI predictions. For example, higher confidence might be assumed for a protein pair where both proteins occur in the same location and share the same GO term for a cellular process. This information is presented in Appendix 4-3 and can be used to form a priority list of interactions for further biological analysis.

Table 4-1: Percentages of *Homo sapiens* pairs in which both partners share the same GO, gene ontology, annotation as well as third party interactions. * P-value $\leq 1\text{E-}16$

	Derived from GO annotation				Third Party Interaction
	Cellular Component (CC)	Molecular Function (MF)	Biological Process (BP)	CC & MF & BP	
(a) Random <i>H. sapiens</i> pairs	19.7%	8.2%	2.8%	0.8%	0.4%
(b) Previously reported <i>H. sapiens</i> interactions	77.2%	64.4%	46.6%	35.0%	59.2%
(c) Predicted <i>H. sapiens</i> interactions identified in this study	64.1%*	43.9%*	30.8%*	20.6%*	23.9%*

Figure 4-1: Distribution of the interacting protein pairs on the basis of subcellular localization (A), molecular function (B) and cellular process (C) for both previously detected interactions and interactions unique to this study normalized by the number of possible pairs with both GO terms. The overall co-occurrence (association) for pairs that were previously not reported is similar to that of previously reported interactions. The observed enrichment for certain categories within predicted interactions may represent new association or cross-communication. For example, in panel B, a co-occurrence (association) for “metal ion binding” and “transcription” is observed for predictions that were not previously reported. Similarly, in panel C, “immune response” and “signal transduction” have a more profound association among the predicted interacting pairs in this study.

In addition to the above cross-validation, we independently evaluated the competence of our predictions by evaluating experimental data gathered using the Lentivirus-delivered, Gateway-compatible affinity tagging system coupled with mass spectrometry (LGTS-MS) approach that has been recently developed for the identification of PPIs in mammalian cell lines (Mak *et al.*, 2010). The LGTS system uses a versatile affinity (VA)-tag constructed in-frame with a Gateway cassette consisting of 3x Flag, 6x His, and 2x Streptactin (Strep) epitopes, with flag and His separated by dual TEV protease cleavage sites for efficient affinity purification (Mak *et al.*, 2010). Using this approach, we stably expressed four (CBX1 (P83916), RNF2 (Q99496), H2AFX (P16104), and RBBP4 (Q09028)) C-terminal affinity tagged chromatin-related proteins in human embryonic kidney (HEK) 293 cells that play an important role in the epigenetic control of chromatin structure and gene expression, transcriptional repression, nucleosome remodeling, and chromatin assembly (Vogel *et al.*, 2006; Sanchez *et al.*, 2007; Kim *et al.*, 2009; Rao *et al.*, 2009; Margueron and Reinberg, 2010) (Figure 4-2A). These tagged proteins were affinity-purified in one step on anti-FLAG resins and the interacting proteins were identified by tandem mass spectrometry. As expected, we recovered both the bait and several well-known interacting protein partners, such as the interaction between the tagged histone binding protein, RBBP4 (Q09028) and subunits of the core histone deacetylase complex (HDAC1 (Q13547) and RBBP7 (Q16576)), confirming the overall efficacy of the protein purification procedure employed in the identification of co-purifying interacting proteins (Figure 4-2B).

Figure 4-2: Affinity purification experiments using P83916 (CBX1), Q99496 (RNF2), P16104 (H2AFX), and Q09028 (RBBP4) as baits. (A) Immunoblot confirming the expression of the indicated FLAG-tagged chromatin proteins using antibody against the 3X FLAG epitope. Two independent FLAG tag constructs of each chromatin related protein was constructed for affinity purifications to eliminate background contaminants and to uncover highly reproducible interactions. Molecular masses (kDa) of marker proteins by SDS-PAGE are indicated. (B) Representative functional clusters identified from affinity purification data (light gray) and expanded by MP-PIPE predictions (dark gray). Tagged-baits are shown by yellow nodes (ellipses). Blue nodes represent co-purifying proteins identified through affinity purification. Purple nodes represent proteins added to the functional clusters through MP-PIPE predictions. Red dashed edges (lines) represent previously reported binary interactions identified in literature experimental data and green solid edges represent novel (not previously reported) MP-PIPE binary interaction predictions.

Consistent with the biological expectation, these co-purifying proteins were enriched for more chromatin related functions such as chromatin organization, binding and assembly, nucleosome assembly, histone ubiquitination and transcriptional regulation. Examples of functional clusters identified through the LGTS-MS based method are shown in Figure 4-2B. To investigate MP-PIPE's ability to explain the observed LGTS-MS data, we computed the precision and recall of MP-PIPE predicted interactions. This is summarized in Table 4-2 where reachable proteins are defined as those proteins that interact directly or through one or two intermediary proteins. This accounts for the fact that a bait and prey observed to co-purify in a LGTS-MS experiment may, in fact, interact indirectly through one or more intermediary proteins. For the four baits, previously known PPI interactions (high confidence literature data) can only explain on average 10.89% (recall) of the co-purifying proteins (prey). Using MP-PIPE predictions increases our recall by ~3-fold (29.31%) while maintaining comparable precisions.

Our predicted interactions appear to have a wide coverage of the human proteome. For instance, of the 22,513 human potential open reading frames included in this study, 11,194 were found in our prediction list (i.e., form at least one interaction) for a coverage of approximately 50%. Since a total of 172,132 interactions were predicted, on average there appears to be approximately 15 interactions for each protein found in the prediction list. As illustrated in Figure 4-3, approximately 32% of the predicted interactions occur in the nucleus, followed by 21% in the cytoplasm. In fact, the distribution of the identified interactions is very consistent with those of the previously reported (known ones). This distribution is also in accordance with the previously reported PPI distribution in *S. cerevisiae* (Krogan *et al.*, 2006). Of interest are membrane proteins, which although not

readily amenable to experimental assays, received good coverage using MP-PIPE. The full range of biological processes and molecular functions are also well covered (Appendix 4-4). On the basis of expression level (obtained from ArrayExpress EMBL-EBI), 66.26% and 68.99% of highly and low expressed proteins, respectively, also appear in the list of PPIs. We also examined the interactions for fibroblast growth factors (FGFs) and cyclin-dependent kinases (CDKs) with fibroblast growth factor receptors (FGFRs) and regulatory inhibitors and activators of cyclin-dependent kinases, respectively. These represent examples of proteins that share high similarity in primary sequence and molecular function, yet have differences in substrate specificity and regulatory factors. Shown in Appendix 4-5 there are clear differences in interacting partners for different members of FGF and CDK proteins. Altogether these observations suggest that our prediction method appears to be inclusive and specific, and is amenable to diverse set of proteins presenting a good coverage of the proteome.

Table 4-2: Overlap of co-purifying proteins identified through LGTS-MS with previously reported (known) interactions and MP-PIPE predictions.

¹Recall calculated as reached/# of prey

²Precision calculated as reached/reachable

		Reachable proteins		Prey reached		Recall ¹		Precision ²	
Bait	# of prey	Known	MP-PIPE	Known	MP-PIPE	Known	MP-PIPE	Known	MP-PIPE
Q09028	301	112	201	56	99	18.60%	32.89%	50.00%	49.25%
P83916	474	91	244	59	178	12.45%	37.55%	64.84%	72.95%
P16104	209	39	207	11	82	5.26%	39.23%	28.21%	39.61%
Q99496	292	16	24	13	15	4.45%	5.14%	81.25%	62.50%
Total	1276	258	676	139	374	10.89%	29.31%	53.88%	55.33%

Figure 4-3: A) Number and B) percentage of co-localized interacting pairs by GO components in previously reported data compared with those reported in this study only. Note that since proteins can have multiple tags, interacting pairs can be co-localized in several components and can be counted more than once.

4.4.4 The proteome-wide PPI network can identify translation genes

The process of protein synthesis or translation is the process by which the genetic messages embedded in mRNAs is sequentially read and converted into polypeptide sequences. Due to its absolute requirement for the survival of a cell, this process has remained highly conserved through the course of evolution. Our predicted interactome included novel interactions for five human proteins Q96DG6 (CMBL), Q08AM6 (VAC14), P23511 (NFYA), Q9UKR5 (ERG28) and P48735 (IDH2) with proteins known to play roles in the process of translation. To study the involvement of these five proteins in translation, we subjected their corresponding yeast homologs (AIM2, VAC14, HAP2, ERG28 and LYS12, respectively) to experimental analysis. First, we examined the effect of their deletion on stop-codon read-through using three different expression plasmids, pUKC817, pUKC818 and pUKC819 that carry premature stop codons UAA, UGA and UAG, respectively, within a *β-galactosidase* reporter gene (expression cassette). As evident from an increase in relative *β-galactosidase* activity shown in Figure 4-4A; the deletion of HAP2, ERG28 and LYS12 have significantly altered the ability of ribosomes to detect all three stop codons. To confirm that the observed elevation of *β-galactosidase* was at the translation level, mRNA content of *β-galactosidase* was measured. No difference between relative content of *β-galactosidase* mRNAs was observed in deletion and control strains (Figure 4-4B). We then further investigated the involvement of HAP2, ERG28 and LYS12 in translation by subjecting their deletion mutants to drugs that affect translation. *hap2Δ*, *erg28Δ* and *lys12Δ* showed altered levels of sensitivity to streptomycin and cycloheximide (Figure 4-4C). Next, translation efficiency (rate) was measured using an inducible LacZ expression cassette on a p416 plasmid. Deletion

mutants for ERG28 and LYS12 had a drastic reduction in the rate of induced LacZ synthesis further linking ERG28 and LYS12 to translation (Figure 4-4D).

Interestingly ERG28 is a well-characterized protein involved in ergosterol biosynthetic pathway, the relation of which to translation is not readily expected. However, in agreement with a link to translation, ERG28 was previously shown to physically interact with a polysome associated mRNA binding protein SLF1 (Schenk *et al.*, 2012), and a putative RNA helicase SPB4 that sediments with 66S pre-ribosomes (Garcia-Gomez *et al.*, 2011). Further, ERG28 is localized to ER membranes and a general link between sterol biosynthesis and translation has previously been proposed (Benko *et al.*, 2000).

Figure 4-4: A) The relative *β-galactosidase* activity is determined by normalizing the activity of the mutant strains carrying different stop-codon read through cassettes to the control construct (pUKC815 with no premature stop codon) in the wild type strain. B) The relative mRNA level is determined by normalizing the mRNA content of the mutant strains carrying different premature stop-codon expression cassettes to those in the wild type. C) Increased sensitivity of *hap2Δ*, *erg28Δ* and *lys12Δ* to different translation inhibitory drugs. Sensitivity of the wild type strain was used as a point of reference. Sensitivity was quantified as low, moderate and high with respect to that for the wild type strain. *hap2Δ*, *lys12Δ*, and *erg28Δ* show increased sensitivity to one or both streptomycin and/or cycloheximide. D) Effect of gene deletions on translation efficiency. Relative translation efficiency was measured using p416 plasmid containing Gal-inducible promoter in LacZ expression cassette normalized to mRNA content. Values are related to translation efficiency of the control strain set at 1.

4.4.5 Identification of novel molecular markers for seasonal allergic rhinitis

Glucocorticoids (GCs) have a key role in the treatment of patients with seasonal allergic rhinitis (SAR) and other allergic disorders (Bousquet *et al.*, 2010). Because of this and difficulties in evaluating treatment response based on clinical signs and symptoms, there is a need for protein markers to monitor that response. The identification of such markers is complicated by the involvement of a large number of inflammatory proteins in SAR (Bousquet *et al.*, 2008). We hypothesized that novel biomarkers could be identified among proteins predicted to interact with proteins belonging to known inflammatory pathways in SAR including the acute phase response pathway, complement signaling pathway and glucocorticoids receptor pathway (Wang *et al.*, 2011a; 2011b).

Proteins from the acute phase response pathway, complement signaling pathway and glucocorticoids receptor pathway were extracted from the Ingenuity pathway analysis (IPA) software. Interactors of these proteins were selected from our predicted human PPI network. We included secreted, membrane and cytoplasmic proteins, but excluded nuclear proteins. We prioritized candidate biomarkers based on their number of known and predicted interactions with proteins known to be involved in SAR-associated responses. Next, we focused on proteins with a high number of predicted interactions.

From the literature we extracted 191 proteins that belong to the acute phase response pathway, complement pathway, and glucocorticoids receptor pathway (Appendix 4-6). These proteins formed the known set of SAR-associated proteins (SARp). From our predicted human PPIs, the proteins that interacted with SARp were determined. A total of 3334 proteins were found to interact with one or more SARp. We prioritized five new

proteins with a high number of total and predicted interactions to SARp as candidate biomarkers, namely PRB1, PRB2, SFN, LYN and Akt2. Using ELISA, we analyzed these candidates in nasal fluid from 40 patients with SAR before and after GC treatment. This study represented protein expression analysis for 400 samples (5 proteins, before and after GC treatment, in 40 patients).

It was observed that after GC treatment LYN concentration increased from 396.1 ± 30.5 pg/mL to 537.7 ± 35.5 pg/mL (P-value ≤ 0.001). PRB1 decreased from 16.3 ± 7.0 ug/mL to 8.5 ± 2.1 ug/mL (P-value ≤ 0.05) (Figure 4-5). PRB2 was not differentially expressed before and after treatment, and SFN and Akt2 were not detectable in most samples. Differential protein expression for LYN and PRB1 provides a good evidence for the possibility of using these proteins as novel molecular markers for SAR. Altogether, the data presented here illustrates the suitability of the predicted PPIs for identifying potential new molecular markers for human conditions.

Figure 4-5: Analysis of candidate proteins with ELISA. Proteins were analyzed in nasal fluid from 40 patients with SAR before and after GC treatment. Pre: patients before GC treatment and Post: patients after GC treatment. LYN and PRB1 were differentially expressed before and after GC treatment with P-values of ≤ 0.001 and ≤ 0.05 , respectively.

4.4.6 Network-wide analysis of hubs and betweenness centrality

In a PPI network, the degree of interaction for a target protein is believed to be a good indicator for the biological importance of that protein within the system (Jeong *et al.*, 2001; Yu *et al.*, 2004). Removal of the highly connected proteins or “hubs” appears to have a more profound effect on the integrity of the network by reducing the size of the largest connected module, than removal of random proteins (Albert *et al.*, 2000). We studied the top 50 hubs with the highest number of interactions within our predicted PPI network, and observed very high enrichment for proteins that affect transcription (P-value: 1.93E-14) and gene expression (P-value: 2.82E-14) (Table 4-3). Transcription factors mediate differential genetic programming and hence are of central importance in developmental biology (Young, 2014), responses to stimuli (Yosef and Regev, 2011), disease progression (Bithell *et al.*, 2009), etc. Betweenness centrality is another topological feature of a network and evaluates the number of shortest paths that pass through a given node (Girvan and Newman, 2002). Therefore, high betweenness centrality for a protein represents the relative number of shortest paths that are associated with that protein. Consequently, proteins with high betweenness centrality are thought to play a central role in the cross-talk and communication between interconnected modules of a network by forming “traffic bottlenecks” for communication (Yu *et al.*, 2007a). We evaluated the top 50 proteins with highest betweenness centrality values within our predicted network. Consistent with the expected role of these proteins in signaling, we observed that they were highly enriched for proteins involved in intracellular communication (kinase activity: P-value: 8.61E-18; signaling: P-value: 7.45E-12), or for

which communication is of central importance (regulation of cell death; P-value: 1.82E-12).

Centrality measurements are often used to predict the possible involvement of a protein in disease etiology and progression. Some studies suggest that hubs are likely enriched for disease proteins, whereas others incline towards betweenness centrality as a better indicator (Goh *et al.*, 2007; Feldman *et al.*, 2008; Ozgur *et al.*, 2008; Chen *et al.*, 2009; Dezso *et al.*, 2009). We therefore examined a possible relationship between the top 500 proteins with the highest degrees of centrality (hub and betweenness centrality) and their reported involvement in disease progression. As illustrated in Figure 4-6, both hub and betweenness centralities appear to be good indicators for disease proteins. However, betweenness centrality appeared to have a better correlation than hubs for disease proteins. A ranked list of top 500 proteins according to their centrality measures is reported in Appendix 4-7 (hubs) and Appendix 4-8 (betweenness centrality). We note that the relationship between connectivity (with hubs having a higher connectivity) and disease is dependent on a variety of factors, particularly gene essentiality (Goh *et al.*, 2007), which we have not investigated here in depth.

Table 4-3: Enrichment of biological process for proteins with highest centrality measurements; hubs and betweenness centrality (top 50).

Biological Process	Number of proteins	P-value
Hubs		
Transcription regulation	28	1.93E-14
Regulation of gene expression	29	2.82E-14
Betweenness Centrality		
Protein kinase activity	20	8.61E-18
Regulation of cell death	19	1.82E-12
Signaling	33	7.45E-12

Figure 4-6: Comparing the number of disease-associated proteins with high degree (hubs) and high betweenness centrality. Proteins with highest betweenness centrality appear to be more enriched for disease proteins than hubs.

4.4.7 Using the predicted human protein interaction network to assign breast cancer proteins

Breast cancer is the most commonly diagnosed form of cancer among women (Yarden and Papa, 2006). BRCA1 and to some extent BRCA2 are the two key genes associated with breast cancer progression. Breast cancer susceptibility has been related to a mutation of BRCA1 (Hall *et al.*, 1990). Carriers of BRCA1 (and some BRCA2) mutations have a 50-80% increased risk of developing breast cancer (Easton *et al.*, 1995). It is estimated that 10% of western women fall in this category (Yarden and Papa, 2006). While the tumor suppression property of BRCA1 is well investigated, the molecular mechanism of its activity in tumor prevention is not fully understood (Wu *et al.*, 2010). Figure 4-7 illustrates a brief overview of the breast cancer pathway where BRCA1 plays a central role. As illustrated, (see Figure 4-7) BRCA1 is directly associated with several cellular processes including chromatin remodeling, DNA damage checkpoint activation, DNA damage sensing, and DNA double stranded break (DSBs) repair. BRCA1 plays an essential role in delaying cell cycle progression by its DNA damage checkpoint activity. ATM phosphorylation of p53, a tumor suppressor protein, is mediated by BRCA1 and, in the presence of DNA damage, delays or arrests G1/S transition (Fabbro *et al.*, 2004). BRCA1 is important during S phase and G2/M checkpoint activation through its regulation of kinase activity of Chk1 (Yu and Chen, 2004). Upon DNA damage, H2AX is phosphorylated by ATM and ATR and recruits MDC and RNF8 to the site of the damage. Subsequently, BRCA1 is translocated to the site of damage by ubiquitination of H2A through RNF8 and Ubc13 (Yarden and Papa, 2006) and interacts with the

Mre11/Rad50/Nbs (MRN) complex that is involved in double stranded DNA break repair (Shrivastav *et al.*, 2008).

Examining the proteins involved in the breast cancer pathway for PPIs, MP-PIPE predicted over 3,000 interactions, 424 of which (161 and 263 known and novel interactions, respectively) directly involve BRCA1 (P38398). Studying these interactions can expand our current understanding of the breast cancer pathway. A number of interesting factors were found to form novel interactions with multiple proteins associated with the breast cancer pathway, including CDK3, AURKB, and SMC1b (see Figure 4-7).

CDK3 is a cyclin-dependent kinase that functions in cell cycle progression and mitosis, and plays an essential role in G1/S transition through its activation of the E2F transcription factor family and G0/G1 transition by Rb phosphorylation (Ren and Rollins, 2004). E2F and Rb play a regulatory role in the transcription of BRCA1 (De Siervi *et al.*, 2010), connecting CDK3 to BRCA1. In further agreement with the observed interactions, CDK3 has high expression levels in cancer cells and participates in cell proliferation and transformation by enhancement of ATF1 activity, a gene that physically interacts with BRCA1 (Houvras *et al.*, 2000; Zheng *et al.*, 2008). Furthermore, both BRCA1 and CDK3 are involved in cell cycle transition, further supporting a potential role for CDK3 in breast cancer.

Figure 4-7: Schematic diagram of breast cancer pathway. BRCA1 plays a central role in breast cancer by connecting DNA damage and chromatin remodeling to downstream processes such as cell cycle progression and DNA repair. Black lines (edges) represent biochemical pathways, red and green edges are novel and known PPIs, respectively. Ovals and clouds represent proteins and protein complexes, respectively. Red ovals represent novel proteins that are associated with breast cancer pathway on the basis of the predicted interactions they form.

AURKB has several functions during mitosis, including spindle assembly, chromosome segregation, and cytokinesis (Carmena *et al.*, 2009). AURKB has high sequence similarity with AURKA, another protein of the Aurora kinase family; however they are reported to differ functionally from each other during mitosis (Hans *et al.*, 2009). It is shown that BRCA1 may be phosphorylated by AURKA, resulting in impaired function of BRCA1 in G2/M transition (Ouchi *et al.*, 2004). AURKB has a single reported interaction within breast cancer pathway through BRCC complex (Ryser *et al.*, 2009). The interactions identified here add credibility to the involvement of this protein in breast cancer pathway.

SMC1B is a meiosis-specific protein involved in chromosome segregation during anaphase, synapsis, and recombination (Revenkova *et al.*, 2004). SMC1B is also part of the cohesin complex, which includes SMC1, SMC3, RAD21 and several other proteins (Peters *et al.*, 2008). The cohesin complex plays a role in several cellular processes such as DNA repair, gene expression regulation and chromosome segregation (Peters *et al.*, 2008; Dorsett, 2011). Recent studies showed that several subunits of the cohesin complex are also important in DNA damage response (Dorsett, 2011). In addition, SMC1b has been linked to neck and head cancer (Zhang *et al.*, 2010b), further supporting a role for this protein in cancer.

We also examined the PPI network for mutations associated with resistance to breast cancer therapeutics doxorubicin and Trastuzumab. Individual mutant proteins were analyzed against the human proteome (one-against-all) for their PPIs at a recall of 23% at a precision of 82.1%. In this way, 5 personalized human PPI networks were predicted,

each differing by a mutation in one gene only. The 5 PPI profiles were compared to that of their corresponding control networks. The list of these mutants is found in supplementary Appendix 4-9. Four mutations, P04637a, P04637b, P04637c and P04637d, in p53 (P04637) protein have been linked to resistance to the chemotherapeutic breast cancer drug, doxorubicin (Aas *et al.*, 1996). The PPI profile for P04637b and P04637d was identical to that of the wild type. However, the other two mutants showed some differences. For example, P04626a and P04626c lost their interactions with the nuclear transcription factor Y, NFYC (Q13952) and the ubiquitin conjugating enzyme E2L3 (P68036) involved in nuclear hormone receptors transcriptional activity, among others. Similarly, a truncated form of HER2 (P04626) is responsible for resistance against HER2-targeted breast cancer therapeutics such as Trastuzumab (Zagozdzon *et al.*, 2011). We observed that truncated-HER2 lost several PPIs including an interaction with a G-protein signaling RGS8 (P57771) which functions as an inhibitor of signal transduction, and an interaction with an early growth response protein ERG1 (P18146) involved in cell differentiation. Of interest, the truncated HER2 formed a new interaction with the tumor suppressor p53 protein. A possible explanation for this novel interaction for the truncated HER2 could be that segments of the deleted region might have physically hindered the availability of the region responsible for an interaction with p53 in the wild type form.

4.4.8 Identification of protein complexes within the human interaction network

Protein complexes can be defined as a group of proteins that interact with each other to form a functional unit. Paracliques (Chesler and Langston, 2006; Langston *et al.*, 2008; Eblen *et al.*, 2009) can be computationally identified as a sub group of proteins within the interaction network with high degree of interconnectivity and may define putative

complexes. Given the size of the human PPI network, prediction of paracliques requires advanced computational approaches to complete a thorough analysis within a reasonable timeframe. We have applied a novel graph theoretic approach to automatically identify paracliques within the network (see methods for details). Our analysis led to a number of interesting predictions. For each paraclique, a statistical analysis of gene ontology (GO) term enrichment was performed (data are presented). For example, Paraclique 1359 is a complex of six proteins with 13 interactions (Figure 4-8A). O00151 (PDLIM1) is a cytoskeletal protein that acts as an adapter to bridge other proteins (like kinases) to the cytoskeleton. P20929 (NEB) is a muscle protein involved in maintaining the structural integrity of sarcomeres and membranes associated with the myofibrils (F-actin stabilization). The rest of the members (P08670 (VIM), P14136 (GFAP), P17661 (DES) and P41219 (PRPH)) are intermediated filament proteins. On the basis of GO enrichment (P-value: $6.5E-07$), one may conclude that the activity of this complex is associated with cytoskeleton and structural integrity of the cell.

Paraclique 1409 is a complex of six proteins with 14 interactions (Figure 4-8B). Q02246 (CNTN2) is involved in cell adhesion and the remaining proteins (O94779 (CNTN5), Q02246 (CNTN2), Q12860 (CNTN1), Q8IWV2 (CNTN4), Q9P232 (CNTN3), and Q9UQ52 (CNTN6)) are involved in cell surface interaction during nervous system development. On the basis of GO enrichment, we can assign this complex to cell adhesion (P-value: $2.2E-10$).

Paraclique 2164 is a complex of five proteins with 10 interactions (Figure 4-8C). Three of its members (P32298 (GRK4), P34947 (GRK5) and P43250 (GRK6)) are G protein-

coupled receptor kinase and the remaining two (Q9NP86 (CABP5) and Q9NZU8 (CABP1)) are calcium-binding proteins. Considering the fact that biological interaction between G-protein coupled receptor and calcium-binding proteins has been widely reported and seems essential in signaling pathways, one may conclude that this complex plays a role in G-protein coupled signaling pathway, a claim which is supported by enriched Gene Ontology term (P-value: 3.75E-08).

Figure 4-8: Schematic representation of the interactions for three putative complexes, paracliques, 1359 (A) 1409 (B) and 2164 (C). A high degree of interconnectivity is observed for these complexes. Red and green edges represent previously reported (known), and novel interactions, respectively.

In this study, we present a comprehensive pair wise analysis and prediction of the entire human PPI network using the principles of short co-occurring polypeptide regions as mediators of PPIs. Through this massive computational analysis, we predict approximately 170,000 PPIs, of which 140,000 have not been reported previously. The distribution of the novel PPIs on the basis of sub-cellular localization, molecular function and biological process are very similar to those of previously reported interactions, highlighting the reliability of our prediction. Our predictions are useful for understanding cellular biology as a whole, with approximately 8,000 protein complexes in our inferred interaction network. Furthermore, specific processes can be successfully interrogated using our new predictions: on the basis of inferred interactions we predict and experimentally confirm novel functions for proteins involved in translation pathway and identify new molecular markers for seasonal allergic rhinitis. Our analysis highlights the usefulness of the predicted PPIs for functional analysis of the human proteome. The speed associated with this approach sets the path for investigating the PPI map for individual humans in a timely fashion. Personal (specific to an individual) PPI maps may improve our knowledge of network and personalized medicine.

5 Chapter: Utilizing yeast genetics to identify novel genes involved in translation-related helicase activity

5.1 Abstract

Translation is one of the most important and highly regulated processes within the gene expression pathway. It defines one of the fundamental steps of life which is to form functional protein products from genetic information. Helicases are essential enzymes which are responsible for unwinding (separating) RNA structures. They are utilized during the translation of mRNAs that contain inhibitory secondary structures. In order to systematically identify genes involved in helicase activity, a plasmid (p281-4) containing 4 inhibitory structures upstream of the LacZ expression cassette and a plasmid (p281) with no inhibitory structure (used as a negative control) were transformed into a yeast non-essential-loss-of-function array (~5000 strains). To systematically identify gene deletion mutants that fail to overcome (melt) secondary structures (hairpin loops), a large-scale *β-galactosidase* lift assay was performed at three different temperatures and followed by a quantitative low-throughput *β-galactosidase* assay. In this manner, 24 potential candidates were identified.

5.2 Introduction

Despite decades of intensive research, there remain numerous uncharacterized genes in the model eukaryotic organism the baker's yeast, *Saccharomyces cerevisiae*. Through the reading of messenger RNA (mRNA) and assembly of corresponding peptide chains, the translation machinery produces millions of different functional proteins that are effectors

in cellular processes. Considering the vital role of translation, it can come as no surprise that there is keen interest in the yeast genetics community in filling holes in our current understanding of the process of translation and discovering functional details of the genes and gene products involved within protein synthesis pathway (Pena-Castillo & Hughes, 2007).

Helicases are essential enzymes (proteins) that move directionally along the nucleic acid phosphodiester backbone, separating double-stranded chains such as DNA-DNA, RNA-RNA or DNA-RNA by breaking hydrogen bonds between nucleotides of different strands. They often use energy derived from nucleoside triphosphate (ATP) hydrolysis to perform their activities. Helicases are involved in different biological pathways, including translation. For example, the DEAD-box protein, eIF4A, is one of the earliest RNA helicases discovered in late 1980s. eIF4A is one of the key components of cap-dependent translation initiation complex. The most likely role of eIF4A is to unwind RNA secondary structure in the 5' un-translated region (UTR) of mRNA to mediate ribosome binding or scanning (Abdelhaleem, 2010).

Helicases have two different conformations, open and closed. Which conformation they take depends on the nucleotide state of a bound ATP (Schutz *et al.*, 2008). The closed conformation is stabilized by ATP and RNA in the nucleotide binding pocket. Hydrolysis of the bound ATP to ADP prompts a change to the open conformation. The mechanical force of this change in conformation is thought to be the driving force used for RNA unwinding (Schutz *et al.*, 2008).

During translation, the unwinding ability of an RNA helicase is necessary for unwinding of any secondary structural elements in the 5' UTR of mRNA prior to the binding of the small ribosomal subunit as well as for proper scanning for the AUG start codon (Iost *et al.*, 1999). Increased complexity of secondary structure, such as dsRNA hairpin loops, increases the need for helicase and proper helicase functioning. It has been shown that insertion of hairpin segments into the 5' UTR, inhibits translation initiation in yeast (Altman *et al.*, 1993). As such, mRNA with increased numbers of hairpins/secondary structures before the AUG start codon are often translated slower than the same mRNA with less secondary structural elements (Altman *et al.*, 1993; Iost *et al.*, 1999).

In this study we used the yeast non-essential gene deletion (known as loss-of-function) collection to identify novel genes involved in translation-related helicase activity. This was done at three different temperatures using a plasmid-based LacZ expression system containing 4 hairpin inhibitory constructs). To this end, 24 candidates were identified that, once deleted, decreased the synthesis of full-length protein directed from an mRNA with hairpin inhibitory structures at 5' end. Many of these genes have unknown functions.

5.3 Materials and methods

5.3.1 Media and strains

The yeast deletion set (*Saccharomyces cerevisiae*, mating type “a” (BY4741)) was used for the large-scale investigation. Yeast mating type “ α ” (BY7092) was used for plasmid transformation and as a mating partner for large-scale plasmid transformation (Tong *et al.*, 2001). DH5 α strain of *Escherichia coli* was used to propagate plasmids (Taylor *et al.*,

1993). Standard rich (YPD) and Synthetic Complete (SC) media were used as growth media for yeast. LB (Lysogeny Broth) was used to grow *E. coli*. Antibiotics were used in the following concentrations: Geneticin, G418, (200 mg/ml) and Ampicillin (50 mg/ml) for selective growth media (Alamgir *et al.*, 2008; 2010).

5.3.2 Plasmids

The p281 (used as a control) and p281-4 reporter plasmids are both yeast plasmids containing a GAL promoter followed by a β -galactosidase reporter gene (LacZ expression cassette). P281-4 additionally has 4 stable hairpin loops between the GAL promoter and the reporter cassette. Original p281 plasmid was generated by (Mueller *et al.*, 1987); p281-4 modified hairpin structure was generated by (Altman *et al.*, 1993).

5.3.3 DNA transformation

DNA (plasmid) transformation in yeast mating type α was performed by the Lithium acetate (LiAc) method described by (Schiestl *et al.*, 1989), and plasmid transformations of chemi-competent *E. coli* followed the protocols of (Inoue *et al.*, 1990). Modified version of SGA, a technique to systematically generate double deletions in large-scale experiments to investigate genetic interactions, was used to transfer selected plasmids into the yeast non-essential gene deletion set (~10,000 plasmid transformation).

5.3.4 β -galactosidase assay

Large-scale β -galactosidase assay using X-gal was carried out as described by (Favier *et al.*, 1996, 1997; Serebriiskii *et al.*, 2000). Quantitative β -galactosidase assay was performed using ONPG (O-nitrophenyl- α -D-galactopyranoside) as described by (Stansfield *et al.*, 1995; Krogan *et al.*, 2003).

5.4 Results and discussion

In order to systematically identify novel candidates involved in translation related helicase activity, we have transformed a plasmid, p281-4, which carries four inhibitory secondary structures (extended hairpins) upstream of a LacZ expression cassette along with a plasmid, p281, which has no inhibitory secondary structure (a control) into the yeast non-essential gene deletion array (~5000 strains) for a total of approximately 10,000 strain transformations. Systematic transformation was performed using modified version of SGA method which was originally designed to study genetic interactions in yeast (Tong *et al.*, 2001). SGA is designed based on a theory that ~80% of genes in yeast have another partner in an overlapping pathway which can compensate for its absence. SGA is performed by systematically mating a query gene (gene deletion or plasmid borne) in mating type “ α ” with gene deletion array in mating type “a”. After a few rounds of selection, haploid mating type “a” progeny are selected carrying double mutants (or single mutants with a second gene over-expressed in a plasmid). Here, we have used the ability of this technique to systematically transfer our two plasmids (p281-4-LacZ and p281-LacZ) separately into the yeast haploid gene deletion array (mating type “a”) (Tong *et al.*, 2001).

The arrayed colonies were tested for their ability to produce functional *β -galactosidase* using a large-scale lift assay on the basis of X-gal hydrolysis in three different temperatures (30° C, 24° C and 18° C) in triplicate (Figure 5-1). Different temperatures were used as secondary structures are known to be further stabilized in lower temperatures. This may increase the stringency of our selections. The inhibitory structure located upstream of the LacZ gene would prevent the production of *β -galactosidase*;

unless the inhibitory structure is bypassed (melted). It is expected that deletion of genes that are associated with helicase activities will be less likely to tolerate inhibitory structures during translation and hence will have a lower efficiency of *β -galactosidase* production.

Figure 5-1: A representative yeast gene deletion colony array subjected to X-gal lift assay at 30° C. White colonies (circled in red) represent gene deletion strains which failed to melt (bypass) hairpin inhibitory structure.

To confirm that our large-scale screening correctly identified those colonies that fail to overcome inhibitory structures (to decrease potential false positives), selected colonies with lower efficiency in producing *β -galactosidase* (white colonies) were analyzed quantitatively through ONPG based *β -galactosidase* assay in triplicate (Table 5-1). In this manner we have identified 24 candidate genes that, when deleted, result in decreased helicase activity. Next, we investigated our selected candidates for Gene Ontology (GO) enrichments. As presented in Figure 5-2A, GO enrichment analysis shows that 25% of selected candidates are already known to be associated with helicase activity, which is expected and indicates the good performance on the part of the selectivity of our assay. Approximately 46% of selected candidates have unknown process/function giving us the opportunity to investigate their novel involvement in translation-related helicase activity. Note that ~29% of the selected candidates have other known functions which may indicate the overall connection between different biological pathways (Figure 5-2B); they may also represent a novel secondary function associated with translation helicase activity.

Table: 5-1: Normalized β -galactosidase assay analysis. All the 24 selected candidates' average β -galactosidase activities from at least 3 independent experiments are normalized to the activity of the *WT* and control plasmid. In this manner, the activity of *WT* is assigned 1 and the activities of the selected mutants are compared to this value. With the (P-value ≤ 0.05 all gene deletion strains showed significant reduction in β -galactosidase activity.

Deletion strains	Normalized B-gal	stdev
<i>WT</i>	1	0.0021
<i>bsc6Δ</i>	0.001300428	0.000448896
<i>bud6Δ</i>	0.43692027	0.017618306
<i>dbp3Δ</i>	0.000232047	0.000003611
<i>dbp7Δ</i>	0.000562933	0.000145046
<i>dbp11Δ</i>	0.004264752	0.001018931
<i>eno1Δ</i>	2.07118E-05	2.05051E-06
<i>hal1Δ</i>	2.4896E-05	2.94243E-06
<i>hog1Δ</i>	0.006683697	0.000115658
<i>hpr5Δ</i>	0.004913714	0.000849215
<i>mms22Δ</i>	0.498362148	0.011550584
<i>mus81Δ</i>	0.000315282	7.34656E-07
<i>nam7Δ</i>	0.523259442	0.012001297
<i>pbs2Δ</i>	1.36124E-05	2.17705E-08
<i>pex11Δ</i>	8.68451E-06	3.83771E-09
<i>pus2Δ</i>	3.40096E-05	5.87202E-07
<i>rim20Δ</i>	3.12654E-06	9.47801E-10
<i>syt1Δ</i>	7.58782E-06	1.75426E-09
<i>ygr122c-aΔ</i>	0.425394587	0.004719815
<i>ygr122wΔ</i>	0.000211633	0.000002408
<i>yjl213wΔ</i>	3.83849E-05	7.7669E-09
<i>yol036wΔ</i>	0.004130835	0.000304909
<i>ypr089wΔ</i>	1.30518E-05	5.80546E-11
<i>ypr096cΔ</i>	1.017E-05	6.90822E-09
<i>yrf1-6Δ</i>	1.20526E-05	4.48846E-07

Figure 5-2: A) Gene Ontology annotation analysis based on the strains with reduced β -galactosidase activity indicated that ~25% of the selected candidates are involved in helicase activities and ~46% of the selected candidates have unknown function or process. B) Interaction map based on merged data (this study and other reported ones) has shown that our selected candidates are enriched for helicase activity (P-value: 8.93E-04) and ATP-dependent RNA helicase activity (P-value: 6.36E-03). Edges represent previously reported association as per color coded.

To have a better coverage (comprehensive analysis), we included the previously reported co-expression, co-localization, genetic interactions (GI) and physical interactions (PPI) data to the list of our 24 positive hits and repeated GO enrichment analysis. In this way we observed that our selected candidates or their common associated partners are enriched for helicase activity (P-value: 8.93E-04) and ATP-dependent RNA helicase activity (P-value: 6.36E-03). More interestingly, five of our selected candidates (*DBP3*, *DBP7*, *NAM7*, *PUS2* and *YGR122W*) are known to be involved in the translation pathway, increasing their likelihood of involvement in translation-related helicase activities (Figure 5-2B).

As mentioned before, we are particularly interested in genes with unknown functions. In this way we may increase our overall coverage of gene functions for the entire yeast genome. For example, *YPR096C* is a gene with unknown function that may interact with ribosomes based on co-purification experiments (Fleischer *et al.*, 2006). It has been shown that *YPR096C* has interaction with genes involved in mRNA binding (*NPL3*, *PAP2* and *VMA1*), genes involved in ATPase activities (*PAP2* and *VMA1*) and genes known to be involved in hydrolase activity (*PAP2*, *PPH21*, *SAP30* and *VMA1*) (Costanzo *et al.*, 2010, Moehle *et al.*, 2012). β -galactosidase analysis has indicated that when *YPR096C* is deleted (*ypr096cΔ*), the efficiency of hairpin inhibitory structure meltdown drops dramatically (relative activity of 0.00001 compared to the *WT* strain).

YPR089W is another protein of unknown function. Drug sensitivity analysis has indicated that *ypr089wΔ* is sensitive to translation-inhibiting reagents including cycloheximide and paromomycin (Alamgir *et al.*, 2010). When deleted, cells have a very low β -

galactosidase activity indicating that *ypr089wΔ* mutant has lower efficiency of synthesis of the β -*galactosidase* hydrolase enzyme due to the hairpin inhibitory structure. This is in accord with GI and PPI data which indicates that *YPR089W* interacts with genes involved in RNA binding (*ARC15*, *DBP9*, *SET1*, *HEK2*, *NPL3*, *PUB1*, *RMP1* and *VMA1*) and hydrolase-ATPase (*ARF2*, *CDC50*, *DBP9*, *GLO4*, *HMI1*, *IRC5*, *PTP1*, *SPB4*, *SSB1*, *VMA* complex and *YPT31*) pathways (Hasegawa *et al.*, 2008; Costanzo *et al.*, 2010, Moehle *et al.*, 2012).

Altogether, based on overall interaction data and our β -*galactosidase* assay, we may claim the involvement of *YPR096C* and *YPR089W* in translation-related helicase activities. However, a more detailed investigation to mechanistically characterize their involvement in translation-related helicase activity is needed.

In conclusion, we have identified 24 gene candidates that, when deleted, reduce expression of a reporter gene from a mRNA that carries inhibitory structures. Further work is needed to confirm their involvement in translation-related helicase activities.

Possible experiments which are highly recommended are as following:

- In vitro helicase assay analysis to confirm the helicase activity of selected candidates.
- Gel Shift analysis to investigate the protein-RNA interactions.
- More detailed genetic interactions analysis including SGA, SDL and PSA to identify genes, clusters or specific pathways genetically interacts with our selected candidates.

6 Chapter: Conclusion

6.1 Conclusion

The budding yeast, *Saccharomyces cerevisiae*, has been extensively used as a model organism to investigate various cellular pathways/processes in eukaryotic systems. With the availability of various gene deletion arrays including MATa and MAT α heterozygous (diploid) deletions, MATa and MAT α haploid deletions, and MATa and MAT α diploid deletions, yeast has become a logical choice for functional genomics studies. In this way phenotypes associated with loss of functions can be readily investigated for various genes in a genome-wide manner. One of the challenging issues regarding large-scale (genome-wide) analysis is the massive amount of data generated by these techniques. Organizing and analysis of such “Big Data” presents a significant challenge for biologists (Jessulat *et al.*, 2011). For this purpose various computational tools continue to be developed that help scientist better organize, visualize, understand and extract meaningful information from what often appears as merely “Raw Data” (Memarian *et al.*, 2007; Jessulat *et al.*, 2011). The most significant challenge that biologists are facing however, remains the notable rates of both false positives as well as false negatives that are associated with big data. False positives represent data selected as positive hits that were not true positives and, false negative hits are data that were not selected (as positive hits) in a screen or experiment but are true positives. The rate of false positives/negatives can vary between ~10-45 % for different experiments. The rate of false positives for certain Y2H data can reach as high as ~45-50% (Delneri *et al.*, 2001; Oliver *et al.*, 2002; Phelps *et al.*, 2002; Urbanczyk-Wochniak *et al.*, 2003; Venancio *et al.*, 2010; Jessulat *et al.*, 2011). As a

result, data collected from large-scale and high throughput methods are often not completely reliable. A way to increase the reliability of the data is to have a good estimate of the effectiveness of the method to pick true positives. This can be achieved by running preliminary mini-large-scale analysis in which a reasonable number of positive and negative controls are included. In this manner, the sensitivity and specificity of the technique(s) - which could represent the rate of false positives/negatives - can be investigated and the screening conditions can be adjusted accordingly. Another common approach is to consistently compare and analyze the generated data against other data bases generated by similar or other large-scale techniques that might be available. This is a very common practice in yeast functional genomics and a growing number of data bases for yeast genetic are freely available to the scientific community. Regardless of the efforts put on minimizing rates of false positives, it is extremely important to confirm the data gathered through large-scale investigations using directed follow up small-scale analysis where the rates of false positives and false negatives are often significantly lower. In this way, only the confirmed hits can be subjected to detailed examination often saving time and the cost associated with these experiments. Using this approach we estimated that our large-scale experiment to identify genes that affect stop codon read-through in yeast had a ~15% rate of false positive. This is a very good accomplishment for the field. Estimation of the rate of false negatives however is significantly more challenging as it is often difficult to estimate the number of novel true positives. A positive hit might only be positive under a specific experimental condition and not under other conditions (Babu *et al.*, 2011a; Omid *et al.*, 2014). Therefore it is difficult to simply label a hit false or novel.

Besides lab-based experiments, which are time and resource intensive, bioinformatics-based predictions represent alternative ways to generate big data in biology. Although, in most cases, these tools are quick and less expensive in comparison to traditional lab-based experiments, validation of generated data is an absolute requirement for the usability of these data by biologists. The rate of success in bioinformatics approaches varies between ~30-70 % (Jessulat *et al.*, 2011; Pitre *et al.*, 2012). In the context of the computational tools, sensitivity and specificity are the two key factors that determine the rate of false positives and false negatives of a screen. Sensitivity is a measure of true positives that has been generated, while specificity deals with true negatives. As an example, PIPE has a sensitivity of ~30%, meaning that only ~30 % of the potential interactions are predicted; on the other hand, it also has 99.9% specificity, indicating the confidence that can be given to a hit once it is identified. Sensitivity and specificity have an inverse relationship where one can be improved at the expense of the other. For example, if a PPI prediction tool assumes all possible proteins pairs to form an interaction, then that tool will have a sensitivity of 100% as all true positives are picked as hits. However the specificity associated with this tool will be close to 0% as an overwhelming majority of predicted interactions are false positives. For our investigations, we studied the global PPI map of a human cell using PIPE as a PPI prediction tool. Since the total possible protein pairs in human proteome is close to 260,000,000 pairs, it was important for us to have a low specificity rate to increase confidence in our hits. Such a high specificity rate however comes with a price of relatively low sensitivity of around 30%. This sensitivity means that only 30% of the global space of the human proteome can be covered by PIPE and hence 70% of the

interactions will be missed. Future improvement to this tool could include specific filters that can possibly increase the sensitivity without compromising specificity. Use of 3D structures of proteins could help achieve this goal. Since PIPE functions on the basis of small protein motifs, those motifs that appear hidden inside a 3D structure of a given protein can be eliminated from the analysis as they are most likely unavailable to mediate an interaction. Similarly those motifs that appear on the surface of the proteins become more important for prediction. Such motifs could receive extra points as their physical location on a protein could make them more effective in mediating an interaction. Combining co-evolution data and sequence conservation information can also help improve the sensitivity of PIPE data.

Here we have used a yeast, MATa haploid, non-essential gene deletion set (~5000 single gene deletions) to systematically investigate the involvement of novel genes in the translation pathway (translation fidelity and translation related helicase activities). Translation is one of the most important, complicated and highly regulated pathways within a living organism. While the translation pathway has been extensively investigated over the past few decades, there is a gap of knowledge in understanding the comprehensive mechanisms that govern the control of translation as well as the cross-communication between translation and other cellular processes (Babu *et al.*, 2011b). A growing number of novel factors that affect translation further highlight the existence of additional genes that influence translation. One way to study the effect of novel genes on this pathway is to investigate gene deletion effects, i.e. how the absence of specific genes influences the process of protein synthesis. We have thus investigated several aspects of translation pathway including translation fidelity and translation related helicase activities

via genome-wide screenings, and bioinformatics approaches to identify novel candidates in this pathway.

The fidelity of the translation - the accuracy of the translation process- is one of the fundamentals of biological systems. It ensures the accuracy during the synthesis of new proteins. To identify genes that affect this process we have utilized the yeast genome to identify candidates that once deleted affect translation stop codon read-through. A modified version of the SGA technique (originally designed to investigate genetic interaction) was used to systematically transfer three different plasmids carrying premature stop codons within LacZ expression cassette and a plasmid with no premature stop codon (as control) into a yeast non-essential gene deletion collection (~ 20,000 transformations) followed by X-gal colony lift assay. In this manner, we managed to identify ~80 gene candidates that, when deleted, compromised translation fidelity (resulted in a higher rate of premature stop codon read-through). Using Gene Ontology (GO) annotation analysis, we have identified a number of genes involved in the translation pathway (which was expected), genes involved in pathways other than translation (cross-communication between different pathways) and a number of previously uncharacterized genes. Detecting genes involved in translation pathways indicated that the experiment is working properly and those candidates can be used as positive controls. We are mostly interested in genes involved in any pathway other than translation and especially uncharacterized genes so we are able to assign them a new function in the protein synthesis pathway. It should be noted that this screen is by no means exhaustive. We anticipate that there still exist additional factors that affect stop codon read-through. Some of these factors can affect stop codon read-through to a lesser

degree making them more difficult to screen by visual inspection (as we did in the current study). Others may simply affect translation under specific conditions that were not tested here. For example, when translation is compromised by the action of a drug which targets translation.

As was mentioned before, one of the most important limitations of large-scale analysis (X-gal lift assay in this case in particular) is the generation of a large amount of data that could contain discrepancies. In order to reconfirm our data and reduce false positive results we applied all the selected candidates to quantitative β -galactosidase assay. As the X-gal lift assay is not a quantitative method, ONPG-based β -galactosidase assay was used as a low-throughput quantitative assay method. Of the 9 randomly selected gene deletions, 7 were confirmed by small scale analysis suggesting an accuracy of ~80% for our large-scale screening.

From the list of selected genes (~80), we focused on 3 candidates (*OLA1*, *BSC2* and *YNL040W*) to investigate in more detail their involvement in the protein synthesis pathway (translation). *OLA1* is a GTPase that has human homolog, *BSC2* is a gene with unknown function that has been bioinformatically predicted to have a leaky sequence and *YNL040W* is a gene with unknown function. Note that none of these three genes has been previously linked to the translation pathway. β -galactosidase analysis has indicated that the deletion of any of these three genes has a severe negative impact on translation fidelity (causing higher error). Next, we hypothesized that if a gene is involved in translation fidelity it might also be involved in translation efficiency (rate of protein synthesis). We used an inducible LacZ reporter system to quantify the translation

efficiency of the selected mutants. *ola1Δ* and *bsc2Δ* showed higher rates of protein synthesis, however *ynl040wΔ* showed a lower rate of protein synthesis in comparison to the *WT*. Note these three candidates are sensitive to translation-inhibiting reagents indicating further evidence of their involvement in translation pathway. We further investigated the activity of these selected candidates by genetic interaction analysis (SGA and PSA). Genetic interaction analysis showed that *OLA1* genetically interacts with genes involved in regulation of translation, amino acid metabolism, ribosome biogenesis and RNA binding, *BSC2* genetically interacts with genes involved in regulation of translation, and *YNL040W* genetically interacts with genes involved in ribosomal biogenesis.

Altogether, we can declare that *OLA1*, *BSC2* and *YNL040W* are involved in translation pathway and we propose name changes of *Translation associated element 5 (TAE5)* for *BSC2* and *TAE6* for *YNL040W*. This study manages to identify novel genes involved in the process of translation. It also provides evidence for the existence of other genes that can influence translation.

Classically, translation can be divided into four main steps: activation, initiation, elongation and termination. Among these four main steps in translation pathway, initiation of translation is the most complicated and regulated step. Generally speaking, in eukaryotes, translation can be divided into cap-dependent and cap-independent initiation. Cap-dependent translation, in most of the cases, is the dominant and most efficient pathway in translation initiation. However, during stress conditions such as apoptosis and viral infection, cap-dependent translation is prohibited and thus initiation of translation should be carried out in an alternative way. An IRES, or Internal Ribosome Entry Site, is

one alternative way cell might use to initiate translation while cap-dependent translation is blocked. Generally, IRESs are highly structured sequences to which the ribosome, in association with other elements called ITAFs, IRES transacting factors, will be recruited to when cap-dependent translation is blocked.

Previous studies have indicated that the IRES-mediated translation could be compromised in mutants that are sensitive to acetic acid (Silva *et al.*, 2013). To this end, we investigated the involvement of 7 candidate genes in IRES-mediated translation pathway with sensitivity to acetic acid: *ATP22* (translation activator), *MRPS16* (mitochondrial ribosomal protein), *PCS60* (oxalate catabolic process and mRNA binding), *RPL36A* (ribosomal protein), *VPS70* (unknown), *YBR063C* (unknown) and *DOM34* (ribosomal subunits disassociation). The activities of these candidates were investigated using 3 different plasmid-based IRES-mediated β -galactosidase reporter systems. β -galactosidase analysis showed that these candidates have lower IRES-mediated translation compared to the *WT*.

Next, we investigated the genetic interactions of these candidates through SGA and PSA. Genetic interaction analysis (SGA and PSA) has indicated the involvement of selected candidates mainly in ribosome biogenesis and regulation of translation. It is a fact that proper biogenesis of the ribosome is one of the most important factors in efficient binding of ribosome onto the mRNA. Any modification in ribosome biogenesis may have an impact on its efficient binding onto the mRNA sequences - in our case mRNA containing IRES sequences-, which may have a negative effect on the process of translation.

Translation pathway and especially initiation of translation is one of the most regulated pathways with in a living organism. In IRES-mediated translation, ITAFs play a key role in ribosomal proper recruitment into the IRES sequences and regulation of translation. All together, we may conclude that our selected candidates, *ATP22*, *MRPS16*, *PCS60*, *RPL36A*, *VPS70*, *YBR063C* and *DOM34*, are in fact involved in IRES-mediated translation and are serving as ITAFs. We observed many GIs between our candidate genes and those involved in ribosome biogenesis. Therefore it is likely that many of our candidate genes affect IRES-mediated translation through influencing ribosome biogenesis. This requires further investigation. Altogether our findings that these 7 genes affect IRES-mediated translation provide further evidence for the existence of other factors that can influence translation control.

Besides laboratory experimentation, bioinformatics approaches are also able to produce considerable biologically meaningful data. A challenging aspect of bioinformatic approaches is to the validation of the data generated by such tools. We have previously developed a bioinformatics tool called PIPE, protein-protein interaction prediction engine, (Pitre *et al.*, 2006) to predict novel protein-protein interactions in a variety of organisms including yeast (Pitre *et al.*, 2008a) and *Caenorhabditis elegans* (Pitre *et al.*, 2012). In order to identify novel genes involved in translation pathways, we investigated human proteins which have novel (not previously reported) interactions with known proteins involved in translation pathway. We then identified those candidates' conserved homologs in yeast (*HAP2*, *ERG28* and *LYS12*) and applied them to translation fidelity, translation efficiency and drug sensitivity (translation-inhibiting reagents) analysis.

Translation fidelity analysis confirmed that *hap2Δ*, *erg28Δ* and *lys12Δ* have higher rates of stop codon bypass compared to the *WT*. Moreover, translation efficiency analysis showed that *erg28Δ* and *lys12Δ* have lower rates of protein synthesis in comparison to the *WT*. Drug sensitivity analysis using translation-inhibiting reagents, streptomycin and cycloheximide, showed high sensitivity for these mutants. All together, these results gave us sufficient evidence that selected candidates (*hap2*, *erg28* and *lys12*) are involved in the translation pathway. Not only this data represents three novel cases of proteins that affect translation, it also serves as further evidence that data gathered from PIPE predictions contain biologically relevant information. To date we have not exhaustively investigated the PPI human map to identify new translation genes. In light of the current study we recommend further investigation into PIPE data for functional genomics.

6.2 Closing remarks and future directions

The translation process represents one of the most fundamental aspects of a living organism. The identification of more than 14 novel genes involved in the process of protein biosynthesis (translation) suggests that there may be many more genes awaiting discovery. It also highlights our gap of knowledge in understanding details of the processes and factors that govern translation. This study leads the way for similar investigations in identifying novel candidates in the translation pathway. These investigations also reaffirm the usage of large-scale analysis in the functional genomics era. Note that this type of screening is not limited to the process of translation and can be applied for any other processes within a living organism.

In accord with the presence of other novel genes that affect translation pathway, our preliminary large-scale attempts (Chapter 5) to identify novel genes that affect translation-related helicase activity has indicated the existence of a number of novel candidate genes that can affect translation initiation. Further work is needed to confirm the involvement of these selected candidates (especially *YPR096C* and *YPR089W*) in translation-related helicase activity. Possible experiments are as following: helicase assay analysis to confirm helicase activity, gel Shift analysis to investigate the protein-RNA interactions and genetic interactions analysis including SGA, SDL and PSA to identify genes, clusters or specific pathways which selected candidates are genetically interacting with.

Conserved homologs of the identified genes can also be investigated for their involvement in translation-related helicase activity in more complex eukaryotes, including human. In general, investigating novel translation-related genes in more complex eukaryotes, for example human, on the basis of the activity of their conserved homolog in yeast provides a suitable opportunity to better understand and investigate the fundamentals of this important process.

Translation is one of the most extensively studied processes in a cell. The fact that we identified a number of novel factors that affect such a well-studied process highlights the likelihood of the presence of other novel factors that may affect different cellular processes. It also highlights our gap of knowledge in the area of functional genomics. The current study underscores the need for further investigations in identifying novel gene functions.

References

- Aas T, Borresen AL, Geisler S, Smith-Sorensen B, Johnsen H, Varhaug JE, Akslen LA, Lonning PE. Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nat Med.* 1996, 2(7):811-814.
- Abdelhaleem M. Helicases: an overview. *Methods Mol Biol.* 2010, 587:1-12.
- Aderem A. Systems biology: its practice and challenges. *Cell.* 2005, 121(4):511-513.
- Ahmed R, Duncan RF. Translational regulation of Hsp90 mRNA. AUG-proximal 5'-untranslated region elements essential for preferential heat shock translation. *J Biol Chem.* 2004, 279(48):49919-49930.
- Alamgir M, Eroukova V, Jessulat M, Xu J, Golshani A. Chemical-genetic profile analysis in yeast suggests that a previously uncharacterized open reading frame, YBR261C, affects protein synthesis. *BMC Genomics.* 2008, 9:583-595.
- Alamgir M, Erukova V, Jessulat M, Azizi A, Golshani A. Chemical-genetic profile analysis of five inhibitory compounds in yeast. *BMC Chem Biol.* 2010, 10:6-20.
- Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature.* 2000, 406(6794):378-382.
- Allam H, Ali N. Initiation factor eIF2-independent mode of c-Src mRNA translation occurs via an internal ribosome entry site. *J Biol Chem.* 2010, 285(8):5713-5725.
- Almeida B, Ohlmeier S, Almeida AJ, Madeo F, Leão C, Rodrigues F, Ludovico P. Yeast protein expression profile during acetic acid-induced apoptosis indicates causal involvement of the TOR pathway. *Proteomics.* 2009, 9(3):720-732.
- Aloy P, Russell RB. InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics.* 2003, 19(1):161-162.
- Altmann M, Müller PP, Wittmer B, Ruchti F, Lanker S, Trachsel H. A *Saccharomyces cerevisiae* homologue of mammalian translation initiation factor 4B contributes to RNA helicase activity. *EMBO J.* 1993, 12(10):3997-4003.
- Andersen CB, Becker T, Blau M, Anand M, Halic M, Balar B, Mielke T, Boesen T, Pedersen JS, Spahn CM, Kinzy TG, Andersen GR, Beckmann R. Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature.* 2006, 443(7112):663-668.

Andreev DE, Dmitriev SE, Terenin IM, Prassolov VS, Merrick WC, Shatsky IN. Differential contribution of the m7G-cap to the 5' end-dependent translation initiation of mammalian mRNAs. *Nucleic Acids Res.* 2009, 37(18):6135-6147.

Anthony B, Carter P, De Benedetti A. Overexpression of the proto-oncogene/translation factor 4E in breast-carcinoma cell lines. *Int J Cancer.* 1996, 65(6):858-863.

Babu M, Díaz-Mejía JJ, Vlasblom J, Gagarinova A, Phanse S, Graham C, Yousif F, Ding H, Xiong X, Nazarians-Armavil A, Alamgir M, Ali M, Pogoutse O, Pe'er A, Arnold R, Michaut M, Parkinson J, Golshani A, Whitfield C, Wodak SJ, Moreno-Hagelsieb G, Greenblatt JF, Emili A. Genetic interaction maps in *Escherichia coli* reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet.* 2011a, 7(11):e1002377.

Babu M, Aoki H, Chowdhury WQ, Gagarinova A, Graham C, Phanse S, Laliberte B, Sunba N, Jessulat M, Golshani A, Emili A, Greenblatt JF, Ganoza MC. Ribosome-dependent ATPase interacts with conserved membrane protein in *Escherichia coli* to modulate protein synthesis and oxidative phosphorylation. *PLoS One.* 2011b, 6(4):e18510.

Bagriantsev S, Liebman S. Modulation of A β 42 low-n oligomerization using a novel yeast reporter system. *BMC Biol.* 2006, 4:32-44.

Benko AL, Vaduva G, Martin NC, Hopper AK. Competition between a sterol biosynthetic enzyme and tRNA modification in addition to changes in the protein synthesis machinery causes altered nonsense suppression. *Proc Natl Acad Sci U S A.* 2000, 97(1):61-66.

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science.* 2011, 334(6062):1524-1529.

Benson M, Strannegard IL, Strannegard O, Wennergren G. Topical steroid treatment of allergic rhinitis decreases nasal fluid TH2 cytokines, eosinophils, eosinophil cationic protein, and IgE but has no significant effect on IFN-gamma, IL-1beta, TNF-alpha, or neutrophils. *J Allergy Clin Immunol.* 2000, 106(2):307-312.

Bithell A, Johnson R, Buckley NJ. Transcriptional dysregulation of coding and non-coding genes in cellular models of Huntington's disease. *Biochem Soc Trans.* 2009, 37(6):1270-1275.

Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nat Rev Genet.* 2007, 8(6):437-449.

Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S. hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol.* 1989, 9(9):3919-3930.

Boube M, Joulia L, Cribbs DL, Bourbon HM. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell.* 2002, 110(2):143-151.

Bousquet J, Khaltaev N, Cruz AA, Denburg J, Fokkens WJ, Togias A, Zuberbier T, Baena-Cagnani CE, Canonica GW, van Weel C, Agache I, Aït-Khaled N, Bachert C, Blaiss MS, Bonini S, Boulet LP, Bousquet PJ, Camargos P, Carlsen KH, Chen Y, Custovic A, Dahl R, Demoly P, Douagui H, Durham SR, van Wijk RG, Kalayci O, Kaliner MA, Kim YY, Kowalski ML, Kuna P, Le LT, Lemiere C, Li J, Lockey RF, Mavale-Manuel S, Meltzer EO, Mohammad Y, Mullol J, Naclerio R, O'Hehir RE, Ohta K, Ouedraogo S, Palkonen S, Papadopoulos N, Passalacqua G, Pawankar R, Popov TA, Rabe KF, Rosado-Pinto J, Scadding GK, Simons FE, Toskala E, Valovirta E, van Cauwenberge P, Wang DY, Wickman M, Yawn BP, Yorgancioglu A, Yusuf OM, Zar H, Annesi-Maesano I, Bateman ED, Ben Kheder A, Boakye DA, Bouchard J, Burney P, Busse WW, Chan-Yeung M, Chavannes NH, Chuchalin A, Dolen WK, Emuzyte R, Grouse L, Humbert M, Jackson C, Johnston SL, Keith PK, Kemp JP, Klossek JM, Larenas-Linnemann D, Lipworth B, Malo JL, Marshall GD, Naspitz C, Nekam K, Niggemann B, Nizankowska-Mogilnicka E, Okamoto Y, Orru MP, Potter P, Price D, Stoloff SW, Vandenplas O, Viegi G, Williams D. Allergic Rhinitis and its Impact on Asthma (ARIA) 2008 update (in collaboration with the World Health Organization, GA(2)LEN and AllerGen). *Allergy.* 2008, 63 Suppl 86:8-160.

Bousquet J, Schünemann HJ, Zuberbier T, Bachert C, Baena-Cagnani CE, Bousquet PJ, Brozek J, Canonica GW, Casale TB, Demoly P, Gerth van Wijk R, Ohta K, Bateman ED, Calderon M, Cruz AA, Dolen WK, Haughney J, Lockey RF, Lötvall J, O'Byrne P, Spranger O, Togias A, Bonini S, Boulet LP, Camargos P, Carlsen KH, Chavannes NH, Delgado L, Durham SR, Fokkens WJ, Fonseca J, Haahtela T, Kalayci O, Kowalski ML, Larenas-Linnemann D, Li J, Mohammad Y, Mullol J, Naclerio R, O'Hehir RE, Papadopoulos N, Passalacqua G, Rabe KF, Pawankar R, Ryan D, Samolinski B, Simons FE, Valovirta E, Yorgancioglu A, Yusuf OM, Agache I, Aït-Khaled N, Annesi-Maesano I, Beghe B, Ben Kheder A, Blaiss MS, Boakye DA, Bouchard J, Burney PG, Busse WW, Chan-Yeung M, Chen Y, Chuchalin AG, Costa DJ, Custovic A, Dahl R, Denburg J, Douagui H, Emuzyte R, Grouse L, Humbert M, Jackson C, Johnston SL, Kaliner MA, Keith PK, Kim YY, Klossek JM, Kuna P, Le LT, Lemiere C, Lipworth B, Mahboub B, Malo JL, Marshall GD, Mavale-Manuel S, Meltzer EO, Morais-Almeida M, Motala C, Naspitz C, Nekam K, Niggemann B, Nizankowska-Mogilnicka E, Okamoto Y, Orru MP, Ouedraogo S, Palkonen S, Popov TA, Price D, Rosado-Pinto J, Scadding GK, Sooronbaev TM, Stoloff SW, Toskala E, van Cauwenberge P, Vandenplas O, van Weel C, Viegi G, Virchow JC, Wang DY, Wickman M, Williams D, Yawn BP, Zar HJ, Zernotti M, Zhong N. Development and implementation of guidelines in allergic rhinitis—an ARIA-GA2LEN paper. *Allergy.* 2010, 65(10):1212-1221.

- Brass N, Heckel D, Sahin U, Pfreundschuh M, Sybrecht GW, Meese E. Translation initiation factor eIF-4gamma is encoded by an amplified gene and induces an immune response in squamous cell lung carcinoma. *Hum Mol Genet.* 1997, 6(1):33-39.
- Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A.* 2005, 102(5):1572-1577.
- Butland G, Krogan NJ, Xu J, Yang WH, Aoki H, Li JS, Krogan N, Menendez J, Cagney G, Kiani GC, Jessulat MG, Datta N, Ivanov I, Abouhaidar MG, Emili A, Greenblatt J, Ganoza MC, Golshani A. Investigating the in vivo activity of the DeaD protein using protein-protein interactions and the translational activity of structured chloramphenicol acetyltransferase mRNAs. *J Cell Biochem.* 2007, 100(3):642-652.
- Caraglia M, Passeggio A, Beninati S, Leardi A, Nicolini L, Improta S, Pinto A, Bianco AR, Tagliaferri P, Abbruzzese A. Interferon alpha2 recombinant and epidermal growth factor modulate proliferation and hypusine synthesis in human epidermoid cancer KB cells. *Biochem J.* 1997, 324 (Pt 3):737-741.
- Carmena M, Ruchaud S, Earnshaw WC. Making the Auroras glow: regulation of Aurora A and B kinase function by interacting proteins. *Curr Opin Cell Biol.* 2009, 21(6):796-805.
- Carr-Schmid A, Durko N, Cavallius J, Merrick WC, Kinzy TG. Mutations in a GTP-binding motif of eukaryotic elongation factor 1A reduce both translational fidelity and the requirement for nucleotide exchange. *J Biol Chem.* 1999, 274(42):30297-30302.
- Casal M, Cardoso H, Leão C. Mechanisms regulating the transport of acetic acid in *Saccharomyces cerevisiae*. *Microbiology.* 1996, 142 (Pt 6):1385-1390.
- Castrillo JI, Oliver SG. Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics. *J Biochem Mol Biol.* 2004, 37(1):93-106.
- Caufield JH, Sakhawalkar N, Uetz P. A comparison and optimization of yeast two-hybrid systems. *Methods.* 2012, 58(4):317-324.
- Chambers A, Tsang JS, Stanway C, Kingsman AJ, Kingsman SM. Transcriptional control of the *Saccharomyces cerevisiae* PGK gene by RAP1. *Mol Cell Biol.* 1989, 9(12):5516-5524.
- Chavatte L, Frolova L, Laugâa P, Kisselev L, Favre A. Stop codons and UGG promote efficient binding of the polypeptide release factor eRF1 to the ribosomal A site. *J Mol Biol.* 2003, 331(4):745-758.
- Chen J, Aronow BJ, Jegga AG. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics.* 2009, 10:73-78.

Chesler EJ, Langston MA. Combinatorial genetic regulatory network analysis tools for high throughput transcriptomic data. In *Systems Biology and Regulatory Genomics*. Volume 4023. Edited by Eskin E: Springer; 2006: 150–165.

Chica C, Diella F, Gibson TJ. Evidence for the concerted evolution between short linear protein motifs and their flanking regions. *PLoS One*. 2009, 4(7):e6052.

Clemens MJ. Targets and mechanisms for the regulation of translation in malignant transformation. *Oncogene*. 2004, 23(18):3180-3188.

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C. The genetic landscape of a cell. *Science*. 2010, 327(5964):425-431.

De Benedetti F, Pignatti P, Bernasconi S, Gerloni V, Matsushima K, Caporali R, Montecucco CM, Sozzani S, Fantini F, Martini A. Interleukin 8 and monocyte chemoattractant protein-1 in patients with juvenile rheumatoid arthritis. Relation to onset types, disease activity, and synovial fluid leukocytes. *J Rheumatol*. 1999, 26(2):425-431.

De Siervi A, De Luca P, Byun JS, Di LJ, Fufa T, Haggerty CM, Vazquez E, Moiola C, Longo DL, Gardner K. Transcriptional autoregulation by BRCA1. *Cancer Res*. 2010, 70(2):532-542.

Delneri D, Brancia FL, Oliver SG. Towards a truly integrative biology through the functional genomics of yeast. *Curr Opin Biotechnol*. 2001, 12(1):87-91.

DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 1997, 278(5338):680-686.

Dezso Z, Nikolsky Y, Nikolskaya T, Miller J, Cherba D, Webb C, Bugrim A. Identifying disease-specific genes based on their topological significance in protein networks. *BMC Syst Biol*. 2009, 3:36-43.

Dorsett D. Cohesin: genomic insights into controlling gene transcription and development. *Curr Opin Genet Dev*. 2011, 21(2):199-206.

Draptchinskaia N, Gustavsson P, Andersson B, Pettersson M, Willig TN, Diansani I, Ball S, Tchernia G, Klar J, Matsson H, Tentler D, Mohandas N, Carlsson B, Dahl N. The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nat Genet*. 1999, 21(2):169-175.

Duttaroy A, Bourbeau D, Wang XL, Wang E. Apoptosis rate can be accelerated or decelerated by overexpression or reduction of the level of elongation factor-1 alpha. *Exp Cell Res*. 1998, 238(1):168-176.

Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Am J Hum Genet*. 1995, 56(1):265-271.

Eblen JD, Gerling IC, Saxton AM, Wu J, Snoddy JR, Langston MA. Graph Algorithms for Integrated Biological Analysis, with Applications to Type 1 Diabetes Data. In *Clustering Challenges in Biological Networks*. Edited by Chaovalitwongse WA: World Scientific; 2009: 207-222.

Ehrenreich IM, Bloom J, Torabi N, Wang X, Jia Y, Kruglyak L. Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet*. 2012, 8(3):e1002570.

Elefsinioti A, Saraç ÖS, Hegele A, Plake C, Hubner NC, Poser I, Sarov M, Hyman A, Mann M, Schroeder M, Stelzl U, Beyer A. Large-scale de novo prediction of physical protein-protein association. *Mol Cell Proteomics*. 2011, 10(11):M111010629.

Espadaler J, Romero-Isart O, Jackson RM, Oliva B. Prediction of protein-protein interactions using distant conservation of sequence patterns and structure relationships. *Bioinformatics*. 2005, 21(16):3360-3368.

Fabbro M, Savage K, Hobson K, Deans AJ, Powell SN, McArthur GA, Khanna KK. BRCA1-BARD1 complexes are required for p53Ser-15 phosphorylation and a G1/S arrest following ionizing radiation-induced DNA damage. *J Biol Chem*. 2004, 279(30):31251-31258.

Fatica A, Tollervey D. Making ribosomes. *Curr Opin Cell Biol*. 2002, 14(3):313-318.

Favier C, Neut C, Mizon C, Cortot A, Colombel JF, Mizon C. Differentiation and identification of human fecal anaerobic-bacteria producing Beta-galactosidase. *Journal of Microbiological Methods*. 1996, 27(1): 25–31.

Favier C, Neut C, Mizon C, Cortot A, Colombel JF, Mizon J. Fecal beta-D-galactosidase production and Bifidobacteria are decreased in Crohn's disease. *Dig Dis Sci*. 1997, 42(4):817-822.

Feldman I, Rzhetsky A, Vitkup D. Network properties of genes harboring inherited disease mutations. *Proc Natl Acad Sci U S A*. 2008, 105(11):4323-4328.

Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature*. 1989, 340(6230):245-246.

Filbin ME, Kieft JS. Toward a structural understanding of IRES RNA function. *Curr Opin Struct Biol*. 2009, 19(3):267-276.

Fleischer TC, Weaver CM, McAfee KJ, Jennings JL, Link AJ. Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes Dev.* 2006, 20(10):1294-1307.

Fromont-Racine M, Senger B, Saveanu C, Fasiolo F. Ribosome assembly in eukaryotes. *Gene.* 2003, 313:17-42.

Fukuchi-Shimogori T, Ishii I, Kashiwagi K, Mashiba H, Ekimoto H, Igarashi K. Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res.* 1997, 57(22):5041-5044.

Gallie DR, Feder JN, Schimke RT, Walbot V. Post-transcriptional regulation in higher eukaryotes: the role of the reporter gene in controlling expression. *Mol Gen Genet.* 1991, 228(1-2):258-264.

Galván Márquez I, Mir-Rashed N, Jessulat M, Atanya M, Golshani A, Durst T, Petit P, Amiguet VT, Boekhout T, Summerbell R, Cruz I, Arnason JT, Smith ML. Antifungal and antioxidant activities of the phytochemistry pipsissewa, *Chimaphila umbellata*. *Phytochemistry.* 2008, 69(3):738-746.

Galván Márquez I, Akuaku J, Cruz I, Cheetham J, Golshani A, Smith ML. Disruption of protein synthesis as antifungal mode of action by chitosan. *Int J Food Microbiol.* 2013, 164(1):108-112.

Galvão FC, Rossi D, Silveira Wda S, Valentini SR, Zanelli CF. The deoxyhypusine synthase mutant *dys1-1* reveals the association of eIF5A and Asc1 with cell wall integrity. *PLoS One.* 2013, 8(4):e60140.

Garcia-Gomez JJ, Lebaron S, Froment C, Monsarrat B, Henry Y, de la Cruz J. Dynamics of the putative RNA helicase Spb4 during ribosome assembly in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 2011, 31(20):4156-4164.

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. Proteome survey reveals modularity of the yeast cell machinery. *Nature.* 2006, 440(7084):631-636.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. Global analysis of protein expression in yeast. *Nature.* 2003, 425(6959):737-741.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P,

Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kötter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*. 2002, 418(6896):387-391.

Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, Chu AM, Jordan MI, Arkin AP, Davis RW. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc Natl Acad Sci U S A*. 2004, 101(3):793-798.

Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*. 2002, 99(12):7821-7826.

Glaser P, Boone C. Beyond the genome: from genomics to systems biology. *Curr Opin Microbiol*. 2004, 7(5):489-491.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. Life with 6000 genes. *Science*. 1996, 274(5287):563-567.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A*. 2007, 104(21):8685-8690.

Gomez SM, Noble WS, Rzhetsky A. Learning to predict protein-protein interactions from protein sequences. *Bioinformatics*. 2003, 19(15):1875-1881.

Goodfellow IG, Roberts LO. Eukaryotic initiation factor 4E. *Int J Biochem Cell Biol*. 2008, 40(12):2675-2680.

Graff JR, Konicek BW, Carter JH, Marcusson EG. Targeting the eukaryotic translation initiation factor 4E for cancer therapy. *Cancer Res*. 2008, 68(3):631-634.

Granneman S, Baserga SJ. Ribosome biogenesis: of knobs and RNA processing. *Exp Cell Res*. 2004, 296(1):43-50.

Grentzmann G, Kelly PJ, Laalami S, Shuda M, Firpo MA, Cenatiempo Y, Kaji A. Release factor RF-3 GTPase activity acts in disassembly of the ribosome termination complex. *RNA*. 1998, 4(8):973-983.

Gubbens J, Kim SJ, Yang Z, Johnson AE, Skach WR. In vitro incorporation of nonnatural amino acids into protein using tRNA(Cys)-derived opal, ochre, and amber suppressor tRNAs. *RNA*. 2010, 16(8):1660-1672.

Guo Y, Yu L, Wen Z, Li M. Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. *Nucleic Acids Res.* 2008, 36(9):3025-3030.

Gursoy A, Keskin O, Nussinov R. Topological properties of protein interaction networks from a structural perspective. *Biochem Soc Trans.* 2008, 36(6):1398-1403.

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science.* 1990, 250(4988):1684-1689.

Han DS, Kim HS, Jang WH, Lee SD, Suh JK. PreSPI: a domain combination based prediction system for protein-protein interaction. *Nucleic Acids Res.* 2004, 32(21):6312-6320.

Hans F, Skoufias DA, Dimitrov S, Margolis RL. Molecular distinctions between Aurora A and B. a single residue change transforms Aurora A into correctly localized and functional Aurora B. *Mol Biol Cell.* 2009, 20(15):3491-3502.

Hartwell LH. Yeast and cancer. *Biosci Rep.* 2004, 24(4-5):523-544.

Hasegawa Y, Irie K, Gerber AP. Distinct roles for Khd1p in the localization and expression of bud-localized mRNAs in yeast. *RNA.* 2008, 14(11):2333-2347.

Hellen CU, Sarnow P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* 2001, 15(13):1593-1612.

Herrera F, Martinez JA, Moreno N, Sadnik I, McLaughlin CS, Feinberg B, Moldave K. Identification of an altered elongation factor in temperature-sensitive mutants 7'-14 of *Saccharomyces cerevisiae*. *J Biol Chem.* 1984, 259(23):14347-14349.

Hieter P, Boguski M. Functional genomics: it's all how you read it. *Science.* 1997, 278(5338):601-602.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science.* 2008, 320(5874):362-365.

Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sørensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature.* 2002, 415(6868):180-183.

Hoepfner D, Helliwell SB, Sadlish H, Schuierer S, Filipuzzi I, Brachat S, Bhullar B, Plikat U, Abraham Y, Altorfer M, Aust T, Baeriswyl L, Cerino R, Chang L, Estoppey D, Eichenberger J, Frederiksen M, Hartmann N, Hohendahl A, Knapp B, Krastel P, Melin N, Nigsch F, Oakeley EJ, Petitjean V, Petersen F, Riedl R, Schmitt EK, Staedtler F, Studer C, Tallarico JA, Wetzel S, Fishman MC, Porter JA, Movva NR. High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. *Microbiol Res*. 2014, 169(2-3):107-120.

Holcik M, Sonenberg N. Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol*. 2005, 6(4):318-327.

Holcik M. Targeting translation for treatment of cancer--a novel role for IRES? *Curr Cancer Drug Targets*. 2004, 4(3):299-311.

Hopper AK. Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics*. 2013, 194(1):43-67.

Houdebine LM, Attal J. Internal ribosome entry sites (IRESs): reality and use. *Transgenic Res*. 1999, 8(3):157-177.

Houvras Y, Benezra M, Zhang H, Manfredi JJ, Weber BL, Licht JD. BRCA1 physically and functionally interacts with ATF1. *J Biol Chem*. 2000, 275(46):36230-36237.

Hsu CH, Wang TY, Chu HT, Kao CY, Chen KC. A quantitative analysis of monochromaticity in genetic interaction networks. *BMC Bioinformatics*. 2011, 13(2):S16.

Huang TW, Tien AC, Huang WS, Lee YC, Peng CL, Tseng HH, Kao CY, Huang CY. POINT: a database for the prediction of protein-protein interactions based on the orthologous interactome. *Bioinformatics*. 2004, 20(17):3273-3276.

Inoue H, Nojima H, Okayama H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene*. 1990, 96(1):23-28.

Iost I, Dreyfus M, Linder P. Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase. *J Biol Chem*. 1999, 274(25):17677-17683.

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A*. 2001, 98(8):4569-4574.

Jackson SE. Hsp90: structure and function. *Top Curr Chem*. 2013, 328:155-240.

Janzen DM, Geballe AP. Modulation of translation termination mechanisms by cis- and trans-acting factors. *Cold Spring Harb Symp Quant Biol*. 2001, 66:459-467.

- Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001, 411(6833):41-42.
- Jessulat M, Alamgir M, Salsali H, Greenblatt J, Xu J, Golshani A. Interacting proteins Rtt109 and Vps75 affect the efficiency of non-homologous end-joining in *Saccharomyces cerevisiae*. *Arch Biochem Biophys*. 2008, 469(2):157-164.
- Jessulat M, Pitre S, Gui Y, Hooshyar M, Omid K, Samanfar B, Tan LH, Alamgir M, Green J, Dehne F, Golshani A. Recent advances in protein-protein interaction prediction: experimental and computational methods. *Expert Opin Drug Discov*. 2011, 6(9):921-935.
- Johnson JM, French SL, Osheim YN, Li M, Hall L, Beyer AL, Smith JS. Rpd3- and spt16-mediated nucleosome assembly and transcriptional regulation on yeast ribosomal DNA genes. *Mol Cell Biol*. 2013, 33(14):2748-2759.
- Jorgensen P, Nishikawa JL, Breikreutz BJ, Tyers M. Systematic identification of pathways that couple cell growth and division in yeast. *Science*. 2002, 297(5580):395-400.
- Kapp LD, Lorsch JR. The molecular mechanics of eukaryotic translation. *Annu Rev Biochem*. 2004, 73:657-704.
- Khan SH, Ahmad F, Ahmad N, Flynn DC, Kumar R. Protein-protein interactions: principles, techniques, and their potential role in new drug development. *J Biomol Struct Dyn*. 2011, 28(6):929-938.
- Kim JA, Haber JE. Chromatin assembly factors Asf1 and CAF-1 have overlapping roles in deactivating the DNA damage checkpoint when DNA repair is complete. *Proc Natl Acad Sci U S A*. 2009, 106(4):1151-1156.
- Kim WK, Park J, Suh JK. Large scale statistical prediction of protein-protein interaction by potentially interacting domain (PID) pair. *Genome Inform*. 2002, 13:42-50.
- King HA, Cobbold LC, Willis AE. The role of IRES trans-acting factors in regulating translation initiation. *Biochem Soc Trans*. 2010, 38(6):1581-1586.
- Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, Scott MS, Gramolini AO, Morris Q, Hallett MT, Rossant J, Hughes TR, Frey B, Emili A. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell*. 2006, 125(1):173-186.
- Kisselev LL, Buckingham RH. Translational termination comes of age. *Trends Biochem Sci*. 2000, 25(11):561-566.
- Koller-Eichhorn R, Marquardt T, Gail R, Wittinghofer A, Kostrewa D, Kutay U, Kambach C. Human OLA1 defines an ATPase subfamily in the Obg family of GTP-binding proteins. *J Biol Chem*. 2007, 282(27):19928-19937.

Komar AA, Lesnik T, Cullin C, Merrick WC, Trachsel H, Altmann M. Internal initiation drives the synthesis of Ure2 protein lacking the prion domain and affects [URE3] propagation in yeast cells. *EMBO J.* 2003, 22(5):1199-1209.

Komar AA, Hatzoglou M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle.* 2011, 10(2):229-240.

Kozak M. New ways of initiating translation in eukaryotes? *Mol Cell Biol.* 2001, 21(6):1899-18907.

Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene.* 2002, 299(1-2):1-34.

Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene.* 2005, 361:13-37.

Kozak M. Some thoughts about translational regulation: forward and backward glances. *J Cell Biochem.* 2007, 102(2):280-290.

Kressler D, Hurt E, Bassler J. Driving ribosome assembly. *Biochim Biophys Acta.* 2010, 1803(6):673-683.

Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C, Shilatifard A, Buratowski S, Greenblatt J. Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol.* 2003, 23(12):4207-4218.

Krogan NJ, Peng WT, Cagney G, Robinson MD, Haw R, Zhong G, Guo X, Zhang X, Canadien V, Richards DP, Beattie BK, Lalev A, Zhang W, Davierwala AP, Mnaimneh S, Starostine A, Tikuisis AP, Grigull J, Datta N, Bray JE, Hughes TR, Emili A, Greenblatt JF. High-definition macromolecular composition of yeast RNA-processing complexes. *Mol Cell.* 2004, 13(2):225-239.

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandhi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature.* 2006, 440(7084):637-643.

Lamphear BJ, Kirchweger R, Skern T, Rhoads RE. Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases.

Implications for cap-dependent and cap-independent translational initiation. *J Biol Chem.* 1995, 270(37):21975-21983.

Langer T, Rosmus S, Fasold H. Intracellular localization of the 90 kDA heat shock protein (HSP90alpha) determined by expression of a EGFP-HSP90alpha-fusion protein in unstressed and heat stressed 3T3 cells. *Cell Biol Int.* 2003, 27(1):47-52.

Langston MA, Perkins AD, Saxton AM, Scharff JA, Voy BH. Innovative Computational Methods for Transcriptomic Data Analysis: A Case Study in the Use of FPT for Practical Algorithm Design and Implementation. *The Computer Journal.* 2008, 51(1):26-38.

Le Quesne JP, Spriggs KA, Bushell M, Willis AE. Dysregulation of protein synthesis and disease. *J Pathol.* 2010, 220(2):140-151.

Lee AY, St Onge RP, Proctor MJ, Wallace IM, Nile AH, Spagnuolo PA, Jitkova Y, Gronda M, Wu Y, Kim MK, Cheung-Ong K, Torres NP, Spear ED, Han MK, Schlecht U, Suresh S, Duby G, Heisler LE, Surendra A, Fung E, Urbanus ML, Gebbia M, Lissina E, Miranda M, Chiang JH, Aparicio AM, Zeghouf M, Davis RW, Cherfils J, Boutry M, Kaiser CA, Cummins CL, Trimble WS, Brown GW, Schimmer AD, Bankaitis VA, Nislow C, Bader GD, Giaever G. Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science.* 2014, 344(6180):208-211.

Lehner B. Modelling genotype-phenotype relationships and human disease with genetic interaction networks. *J Exp Biol.* 2007, 210(9):1559-1566.

Lesage G, Sdicu AM, Ménard P, Shapiro J, Hussein S, Bussey H. Analysis of beta-1,3-glucan assembly in *Saccharomyces cerevisiae* using a synthetic interaction network and altered sensitivity to caspofungin. *Genetics.* 2004, 167(1):35-49.

Lievens S, Lemmens I, Tavernier J. Mammalian two-hybrids come of age. *Trends Biochem Sci.* 2009, 34(11):579-588.

Lucchini G, Hinnebusch AG, Chen C, Fink GR. Positive regulatory interactions of the *HIS4* gene of *Saccharomyces cerevisiae*. *Mol Cell Biol.* 1984, 4(7):1326-1333.

Ludovico P, Madeo F, Silva M. Yeast programmed cell death: an intricate puzzle. *IUBMB Life.* 2005, 57(3):129-135.

Ludovico P, Sousa MJ, Silva MT, Leão C, Côrte-Real M. *Saccharomyces cerevisiae* commits to a programmed cell death process in response to acetic acid. *Microbiology.* 2001, 147(Pt 9):2409-2415.

Luesch H, Wu TY, Ren P, Gray NS, Schultz PG, Supek F. A genome-wide overexpression screen in yeast for small-molecule target identification. *Chem Biol.* 2005, 12(1):55-63.

- Madeo F, Fröhlich E, Ligr M, Grey M, Sigrist SJ, Wolf DH, Fröhlich KU. Oxygen stress: a regulator of apoptosis in yeast. *J Cell Biol.* 1999, 145(4):757-767.
- Mak AB, Ni Z, Hewel JA, Chen GI, Zhong G, Karamboulas K, Blakely K, Smiley S, Marcon E, Roudeva D, Li J, Olsen JB, Wan C, Punna T, Isserlin R, Chetyrkin S, Gingras AC, Emili A, Greenblatt J, Moffat J. A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. *Mol Cell Proteomics.* 2010, 9(5):811-823.
- Malta-Vacas J, Aires C, Costa P, Conde AR, Ramos S, Martins AP, Monteiro C, Brito M. Differential expression of the eukaryotic release factor 3 (eRF3/GSPT1) according to gastric cancer histological types. *J Clin Pathol.* 2005, 58(6):621-625.
- Margueron R, Reinberg D. Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet.* 2010, 11(4):285-296.
- Martin S, Roe D, Faulon JL. Predicting protein-protein interactions using signature products. *Bioinformatics.* 2005, 21(2):218-226.
- McDowall MD, Scott MS, Barton GJ. PIPs: human protein-protein interaction prediction database. *Nucleic Acids Res.* 2009, 37:D651-656.
- Memarian N, Jessulat M, Alirezaie J, Mir-Rashed N, Xu J, Zareie M, Smith M, Golshani A. Colony size measurement of the yeast gene deletion strains for functional genomics. *BMC Bioinformatics.* 2007, 8:117-127.
- Merl J, Jakob S, Ridinger K, Hierlmeier T, Deutzmann R, Milkereit P, Tschochner H. Analysis of ribosome biogenesis factor-modules in yeast cells depleted from pre-ribosomes. *Nucleic Acids Res.* 2010, 38(9):3068-3080.
- Moehle EA, Ryan CJ, Krogan NJ, Kress TL, Guthrie C. The yeast SR-like protein Npl3 links chromatin modification to mRNA processing. *PLoS Genet.* 2012, 8(11):e1003101.
- Moffat J, Sabatini DM. Building mammalian signalling pathways with RNAi screens. *Nat Rev Mol Cell Biol.* 2006, 7(3):177-187.
- Moskalenko SE, Chabelskaya SV, Inge-Vechtormov SG, Philippe M, Zhouravleva GA. Viable nonsense mutants for the essential gene SUP45 of *Saccharomyces cerevisiae*. *BMC Mol Biol.* 2003, 4:2.
- Mueller PP, Harashima S, Hinnebusch AG. A segment of GCN4 mRNA containing the upstream AUG codons confers translational control upon a heterologous yeast transcript. *Proc Natl Acad Sci U S A.* 1987, 84(9):2863-2867.
- Namy O, Duchateau-Nguyen G, Hatin I, Hermann-Le Denmat S, Termier M, Rousset JP. Identification of stop codon readthrough genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2003, 31(9):2289-2296.

- Nazar RN. Ribosomal RNA processing and ribosome biogenesis in eukaryotes. *IUBMB Life*. 2004, 56(8):457-465.
- Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB. Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol*. 2005, 3(12):e405.
- Neduva V, Russell RB. Peptides mediating interaction networks: new leads at last. *Curr Opin Biotechnol*. 2006, 17(5):465-471.
- Ni Z, Olsen JB, Emili A, Greenblatt JF. Identification of mammalian protein complexes by lentiviral-based affinity purification and mass spectrometry. *Methods Mol Biol*. 2011, 781:31-45.
- Nibbe RK, Chowdhury SA, Koyuturk M, Ewing R, Chance MR. Protein-protein interaction networks and subnetworks in the biology of disease. *Wiley Interdiscip Rev Syst Biol Med*. 2011, 3(3):357-367.
- Nierras CR, Warner JR. Protein kinase C enables the regulatory circuit that connects membrane synthesis to ribosome synthesis in *Saccharomyces cerevisiae*. *J Biol Chem*. 1999, 274(19):13235-13241.
- Obrig TG, Culp WJ, McKeehan WL, Hardesty B. The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes. *J Biol Chem*. 1971, 246(1):174-181.
- Ogmen U, Keskin O, Aytuna AS, Nussinov R, Gursoy A. PRISM: protein interactions by structural matching. *Nucleic Acids Res*. 2005, 33:W331-3236.
- Oliver DJ, Nikolau B, Wurtele ES. Functional genomics: high-throughput mRNA, protein, and metabolite analyses. *Metab Eng*. 2002, 4(1):98-106.
- Omidi K, Hooshyar M, Jessulat M, Samanfar B, Sanders M, Burnside D, Pitre S, Schoenrock A, Xu J, Babu M, Golshani A. Phosphatase complex Pph3/Psy2 is involved in regulation of efficient non-homologous end-joining pathway in the yeast *Saccharomyces cerevisiae*. *PLoS One*. 2014, 9(1):e87248.
- Ouchi M, Fujiuchi N, Sasai K, Katayama H, Minamishima YA, Ongusaha PP, Deng C, Sen S, Lee SW, Ouchi T. BRCA1 phosphorylation by Aurora-A in the regulation of G2 to M transition. *J Biol Chem*. 2004, 279(19):19643-19648.
- Outeiro TF, Giorgini F. Yeast as a drug discovery platform in Huntington's and Parkinson's diseases. *Biotechnol J*. 2006, 1(3):258-269.
- Ozgun A, Vu T, Erkan G, Radev DR. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics*. 2008, 24(13):277-285.

Park Y. Critical assessment of sequence-based protein-protein interaction prediction methods that do not require homologous protein sequences. *BMC Bioinformatics*. 2009, 10:419-425.

Parsons AB, Brost RL, Ding H, Li Z, Zhang C, Sheikh B, Brown GW, Kane PM, Hughes TR, Boone C. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol*. 2004, 22(1):62-69.

Parsons AB, Lopez A, Givoni IE, Williams DE, Gray CA, Porter J, Chua G, Sopko R, Brost RL, Ho CH, Wang J, Ketela T, Brenner C, Brill JA, Fernandez GE, Lorenz TC, Payne GS, Ishihara S, Ohya Y, Andrews B, Hughes TR, Frey BJ, Graham TR, Andersen RJ, Boone C. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell*. 2006, 126(3):611-625.

Pena-Castillo L, Hughes TR. Why are there still over 1000 uncharacterized yeast gene? *Genetics*. 2007, 176(1): 7-14.

Pestova TV, Hellen CU. Functions of eukaryotic factors in initiation of translation. *Cold Spring Harb Symp Quant Biol*. 2001, 66:389-396.

Peters JM, Tedeschi A, Schmitz J. The cohesin complex and its roles in chromosome biology. *Genes Dev*. 2008, 22(22):3089-3114.

Petranovic D, Nielsen J. Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol*. 2008, 26(11):584-590.

Petsalaki E, Stark A, Garcia-Urdiales E, Russell RB. Accurate prediction of peptide binding sites on protein surfaces. *PLoS Comput Biol*. 2009, 5(3):e1000335.

Pfaffl MW, Lange IG, Daxenberger A, Meyer HH. Tissue-specific expression pattern of estrogen receptors (ER): quantification of ER alpha and ER beta mRNA with real-time RT-PCR. *APMIS*. 2001, 109(5):345-355.

Phelps TJ, Palumbo AV, Beliaev AS. Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints. *Curr Opin Biotechnol*. 2002, 13(1):20-24.

Pitre S, Dehne F, Chan A, Cheetham J, Duong A, Emili A, Gebbia M, Greenblatt J, Jessulat M, Krogan N, Luo X, Golshani A. PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. *BMC Bioinformatics*. 2006, 7:365-380.

Pitre S, North C, Alamgir M, Jessulat M, Chan A, Luo X, Green JR, Dumontier M, Dehne F, Golshani A. Global investigation of protein-protein interactions in yeast *Saccharomyces cerevisiae* using re-occurring short polypeptide sequences. *Nucleic Acids Res*. 2008a, 36(13):4286-4294.

- Pitre S, Alamgir M, Green JR, Dumontier M, Dehne F, Golshani A. Computational methods for predicting protein-protein interactions. *Adv Biochem Eng Biotechnol.* 2008b, 110:247-267.
- Pitre S, Hooshyar M, Schoenrock A, Samanfar B, Jessulat M, Green JR, Dehne F, Golshani A. Short Co-occurring Polypeptide Regions Can Predict Global Protein Interaction Maps. *Sci Rep.* 2012, 2:239-249.
- Planta RJ, Mager WH. The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Yeast.* 1998, 14(5):471-477.
- Polunovsky VA, Gingras AC, Sonenberg N, Peterson M, Tan A, Rubins JB, Manivel JC, Bitterman PB. Translational control of the antiapoptotic function of Ras. *J Biol Chem.* 2000, 275(32):24776-24780.
- Preiss T, W Hentze M. Starting the protein synthesis machine: eukaryotic translation initiation. *Bioessays.* 2003, 25(12):1201-1211.
- Qi Y, Bar-Joseph Z, Klein-Seetharaman J. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins.* 2006, 63(3):490-500.
- Rajagopala SV, Sikorski P, Caufield JH, Tovchigrechko A, Uetz P. Studying protein complexes by the yeast two-hybrid system. *Methods.* 2012, 58(4):392-9.
- Rao PS, Satelli A, Zhang S, Srivastava SK, Srivenugopal KS, Rao US. RNF2 is the target for phosphorylation by the p38 MAPK and ERK signaling pathways. *Proteomics.* 2009, 9(10):2776-2787.
- Raught B, Gingras AC, James A, Medina D, Sonenberg N, Rosen JM. Expression of a translationally regulated, dominant-negative CCAAT/enhancer-binding protein beta isoform and up-regulation of the eukaryotic translation initiation factor 2alpha are correlated with neoplastic transformation of mammary epithelial cells. *Cancer Res.* 1996, 56(19):4382-4386.
- Ren S, Rollins BJ. Cyclin C/cdk3 promotes Rb-dependent G0 exit. *Cell.* 2004, 117(2):239-251.
- Revenkova E, Eijpe M, Heyting C, Hodges CA, Hunt PA, Liebe B, Scherthan H, Jessberger R. Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. *Nat Cell Biol.* 2004, 6(6):555-562.
- Rhoads RE. Signal transduction pathways that regulate eukaryotic protein synthesis. *J Biol Chem.* 1999, 274(43):30337-30340.

- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol.* 1999, 17(10):1030-1032.
- Rudra D, Warner JR. What better measure than ribosome synthesis? *Genes Dev.* 2004, 18(20):2431-2436.
- Ruffner H, Bauer A, Bouwmeester T. Human protein-protein interaction networks and the value for drug discovery. *Drug Discov Today.* 2007, 12(17-18):709-716.
- Ruggero D, Pandolfi PP. Does the ribosome translate cancer? *Nat Rev Cancer.* 2003, 3(3):179-192.
- Ryser S, Dizin E, Jefford CE, Delaval B, Gagos S, Christodoulidou A, Krause KH, Birnbaum D, Irminger-Finger I. Distinct roles of BARD1 isoforms in mitosis: full-length BARD1 mediates Aurora B degradation, cancer-associated BARD1beta scaffolds Aurora B and BRCA2. *Cancer Res.* 2009, 69(3):1125-1134.
- Sachs AB, Varani G. Eukaryotic translation initiation: there are (at least) two sides to every story. *Nat Struct Biol.* 2000, 7(5):356-361.
- Samanfar B, Omid K, Hooshyar M, Laliberte B, Alamgir M, Seal AJ, Ahmed-Muhsin E, Viteri DF, Said K, Chalabian F, Golshani A, Wainer G, Burnside D, Shostak K, Bugno M, Willmore WG, Smith ML, Golshani A. Large-scale investigation of oxygen response mutants in *Saccharomyces cerevisiae*. *Mol Biosyst.* 2013, 9(6):1351-1359.
- Samanfar B, Tan LH, Shostak K, Chalabian F, Wu Z, Alamgir M, Sunba N, Burnside D, Omid K, Hooshyar M, Galván Márquez I, Jessulat M, Smith ML, Babu M, Azizi A, Golshani A. A global investigation of gene deletion strains that affect premature stop codon bypass in yeast, *Saccharomyces cerevisiae*. *Mol Biosyst.* 2014, 10(4):916-924.
- Sanchez C, Sanchez I, Demmers JA, Rodriguez P, Strouboulis J, Vidal M. Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbx10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. *Mol Cell Proteomics.* 2007, 6(5):820-834.
- Sandbaken MG, Lupisella JA, DiDomenico B, Chakraborty K. Protein synthesis in yeast. Structural and functional analysis of the gene encoding elongation factor 3. *J Biol Chem.* 1990, 265(26):15838-15844.
- Schenk L, Meinel DM, Strasser K, Gerber AP. La-motif-dependent mRNA association with Slf1 promotes copper detoxification in yeast. *RNA.* 2012, 18(3):449-461.
- Scheper GC, van der Knaap MS, Proud CG. Translation matters: protein synthesis defects in inherited disease. *Nat Rev Genet.* 2007, 8(9):711-723.

- Schiestl RH, Gietz RD. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet*. 1989, 16(5-6):339-346.
- Schiffmann R, van der Knaap MS. The latest on leukodystrophies. *Curr Opin Neurol*. 2004, 17(2):187-192.
- Schüler M, Connell SR, Lescoute A, Giesebrecht J, Dabrowski M, Schroeer B, Mielke T, Penczek PA, Westhof E, Spahn CM. Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat Struct Mol Biol*. 2006, 13(12):1092-1096.
- Schutz P, Bumann M, Oberholzer AE, Bieniossek C, Trachsel H, Altmann M, Baumann U. Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A*. 2008, 105(28):9564-9569.
- Serebriiskii IG, Golemis EA. Uses of lacZ to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system. *Anal Biochem*. 2000, 285(1):1-15.
- Shenton D, Smirnova JB, Selley JN, Carroll K, Hubbard SJ, Pavitt GD, Ashe MP, Grant CM. Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *J Biol Chem*. 2006, 281(39):29011-29021.
- Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. *Cell Res*. 2008, 18(1):134-147.
- Silva A, Sampaio-Marques B, Fernandes A, Carreto L, Rodrigues F, Holcik M, Santos MA, Ludovico P. Involvement of yeast HSP90 isoforms in response to stress and cell death induced by acetic acid. *PLoS One*. 2013, 8(8):e71294.
- Silvera D, Arju R, Darvishian F, Levine PH, Zolfaghari L, Goldberg J, Hochman T, Formenti SC, Schneider RJ. Essential role for eIF4GI overexpression in the pathogenesis of inflammatory breast cancer. *Nat Cell Biol*. 2009, 11(7):903-908.
- Skrabanek L, Saini HK, Bader GD, Enright AJ. Computational prediction of protein-protein interactions. *Mol Biotechnol*. 2008, 38(1):1-17.
- Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B. Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell*. 2006, 21(3):319-330.
- Spriggs KA, Stoneley M, Bushell M, Willis AE. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biol Cell*. 2008, 100(1):27-38.
- Stansfield I, Akhmaloka, Tuite MF. A mutant allele of the SUP45 (SAL4) gene of *Saccharomyces cerevisiae* shows temperature-dependent allosuppressor and omnipotent suppressor phenotypes. *Curr Genet*. 1995, 27(5):417-426.

- Stansfield I, Eurwilaichitr L, Akhmaloka, Tuite MF. Depletion in the levels of the release factor eRF1 causes a reduction in the efficiency of translation termination in yeast. *Mol Microbiol.* 1996, 20(6):1135-1143.
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* 2006, 34(2):535-539.
- Stein A, Aloy P. Novel peptide-mediated interactions derived from high-resolution 3-dimensional structures. *PLoS Comput Biol.* 2010, 6(5):e1000789.
- Stelzl U, Wanker EE. The value of high quality protein-protein interaction networks for systems biology. *Curr Opin Chem Biol.* 2006, 10(6):551-558.
- Stoneley M, Willis AE. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene.* 2004, 23(18):3200-3207.
- Suter B, Auerbach D, Stagljar I. Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond. *Biotechniques.* 2006, 40(5):625-644.
- Szilágyi A, Grimm V, Arakaki AK, Skolnick J. Prediction of physical protein-protein interactions. *Phys Biol.* 2005, 2(2):1-16.
- Tafforeau L, Zorbas C, Langhendries JL, Mullineux ST, Stamatopoulou V, Mullier R, Wacheul L, Lafontaine DL. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Mol Cell.* 2013, 51(4):539-551.
- Takahashi N, Yanagida M, Fujiyama S, Hayano T, Isobe T. Proteomic snapshot analyses of preribosomal ribonucleoprotein complexes formed at various stages of ribosome biogenesis in yeast and mammalian cells. *Mass Spectrom Rev.* 2003, 22(5):287-317.
- Taylor RG, Walker DC, McInnes RR. E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.* 1993, 21(7):1677-1678.
- Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJ, Smith LA, Kavoussanakis K, Koessler T, Pharoah PD, Buch S, Schafmayer C, Tepel J, Schreiber S, Völzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary IJ, Starr JM, Tomlinson IP, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston RS, Rennert G, Ballinger D, Rozek L, Gruber SB, Matsuda K, Kidokoro T, Nakamura Y, Zanke BW, Greenwood CM, Rangrej J, Kustra R, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 2008, 40(5):631-637.

- Thakor N, Holcik M. IRES-mediated translation of cellular messenger RNA operates in eIF2 α - independent manner during stress. *Nucleic Acids Res.* 2012, 40(2):541-552.
- Thompson SR. So you want to know if your message has an IRES? *Wiley Interdiscip Rev RNA.* 2012, 3(5):697-705.
- Thornton S, Anand N, Purcell D, Lee J. Not just for housekeeping: protein initiation and elongation factors in cell growth and tumorigenesis. *J Mol Med (Berl).* 2003, 81(9):536-548.
- Tischler J, Lehner B, Fraser AG. Evolutionary plasticity of genetic interaction networks. *Nat Genet.* 2008, 40(4):390-391.
- Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science.* 2001, 294(5550):2364-2368.
- Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Ménard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C. Global mapping of the yeast genetic interaction network. *Science.* 2004, 303(5659):808-813.
- Triana-Alonso FJ, Chakraborty K, Nierhaus KH. The elongation factor 3 unique in higher fungi and essential for protein biosynthesis is an E site factor. *J Biol Chem.* 1995, 270(35):20473-20478.
- Tschochner H, Hurt E. Pre-ribosomes on the road from the nucleolus to the cytoplasm. *Trends Cell Biol.* 2003, 13(5):255-263.
- Tuite MF, McLaughlin CS. The effects of paromomycin on the fidelity of translation in a yeast cell-free system. *Biochim Biophys Acta.* 1984, 783(2):166-170.
- Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature.* 2000, 403(6770):623-627.
- Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.* 2003, 4(10):989-993.

Vasconcelles MJ, Jiang Y, McDaid K, Gilooly L, Wretzel S, Porter DL, Martin CE, Goldberg MA. Identification and characterization of a low oxygen response element involved in the hypoxic induction of a family of *Saccharomyces cerevisiae* genes. Implications for the conservation of oxygen sensing in eukaryotes. *J Biol Chem*. 2001, 276(17):14374-14384.

Venancio TM, Balaji S, Aravind L. High-confidence mapping of chemical compounds and protein complexes reveals novel aspects of chemical stress response in yeast. *Mol Biosyst*. 2010, 6(1):175-181.

Venema J, Tollervey D. Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet*. 1999, 33:261-311.

Vogel MJ, Guelen L, de Wit E, Peric-Hupkes D, Loden M, Talhout W, Feenstra M, Abbas B, Classen AK, van Steensel B. Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res*. 2006, 16(12):1493-1504.

Von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*. 2002, 417(6887):399-403.

Wang H, Chavali S, Mobini R, Muraro A, Barbon F, Boldrin D, Aberg N, Benson M. A pathway-based approach to find novel markers of local glucocorticoid treatment in intermittent allergic rhinitis. *Allergy*. 2011a, 66(1):132-140.

Wang H, Gottfries J, Barrenäs F, Benson M. Identification of Novel Biomarkers in Seasonal Allergic Rhinitis by Combining Proteomic, Multivariate and Pathway Analysis. *PLoS One*. 2011b, 6(8):e23563.

West MJ, Sullivan NF, Willis AE. Translational upregulation of the c-myc oncogene in Bloom's syndrome cell lines. *Oncogene*. 1995, 11(12):2515-2524.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Véronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*. 1999, 285(5429):901-906.

Wu J, Lu LY, Yu X. The role of BRCA1 in DNA damage response. *Protein Cell*. 2010, 1(2):117-123.

- Wuster A, Babu MM. Conservation and evolutionary dynamics of the agr cell-to-cell communication system across firmicutes. *J Bacteriol.* 2008, 190(2):743-746.
- Wyrick JJ, Holstege FC, Jennings EG, Causton HC, Shore D, Grunstein M, Lander ES, Young RA. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature.* 1999, 402(6760):418-421.
- Yarden RI, Papa MZ. BRCA1 at the crossroad of multiple cellular pathways: approaches for therapeutic interventions. *Mol Cancer Ther.* 2006, 5(6):1396-1404.
- Yosef N, Regev A. Impulse control: temporal dynamics in gene transcription. *Cell.* 2011, 144(6):886-896.
- Young RA. Control of the embryonic stem cell state. *Cell.* 2011, 144(6):940-954.
- Yu CY, Chou LC, Chang DT. Predicting protein-protein interactions in unbalanced data using the primary structure of proteins. *BMC Bioinformatics.* 2010, 11:167-172.
- Yu H, Greenbaum D, Xin Lu H, Zhu X, Gerstein M. Genomic analysis of essentiality within protein networks. *Trends Genet.* 2004, 20(6):227-231.
- Yu X, Chen J. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. *Mol Cell Biol.* 2004, 24(21):9478-9486.
- Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol.* 2007a, 3(4):e59.
- Yu S, Vincent A, Opriessnig T, Carpenter S, Kitikoon P, Halbur PG, Thacker E. Quantification of PCV2 capsid transcript in peripheral blood mononuclear cells (PBMCs) in vitro. *Vet Microbiol.* 2007b, 123(1-3):34-42.
- Zagozdzon R, Gallagher WM, Crown J. Truncated HER2: implications for HER2-targeted therapeutics. *Drug Discov Today.* 2011, 16(17-18):810-816.
- Zhang QC, Petrey D, Norel R, Honig BH. Protein interface conservation across structure space. *Proc Natl Acad Sci U S A.* 2010a, 107(24):10896-10901.
- Zhang X, Yang H, Lee JJ, Kim E, Lippman SM, Khuri FR, Spitz MR, Lotan R, Hong WK, Wu X. MicroRNA-related genetic variations as predictors for risk of second primary tumor and/or recurrence in patients with early-stage head and neck cancer. *Carcinogenesis.* 2010b, 31(12):2118-2123.
- Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T, Maniatis T, Califano A, Honig B. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature.* 2012, 490(7421):556-560.

Zheng D, Cho YY, Lau AT, Zhang J, Ma WY, Bode AM, Dong Z. Cyclin-dependent kinase 3-mediated activating transcription factor 1 phosphorylation enhances cell transformation. *Cancer Res.* 2008, 68(18):7650-7660.

Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. Global analysis of protein activities using proteome chips. *Science.* 2001, 293(5537):2101-2105.

Zhu J, Korostelev A, Costantino DA, Donohue JP, Noller HF, Kieft JS. Crystal structures of complexes containing domains from two viral internal ribosome entry site (IRES) RNAs bound to the 70S ribosome. *Proc Natl Acad Sci U S A.* 2011, 108(5):1839-1844.

Zimmer SG, DeBenedetti A, Graff JR. Translational control of malignancy: the mRNA cap-binding protein, eIF-4E, as a central regulator of tumor formation, growth, invasion and metastasis. *Anticancer Res.* 2000, 20(3A):1343-1351.

Appendices

Part A

Appendix 1-1: A list of available yeast resources (database) derived from large-scale studies.

List of yeast resources	Web address
Array Express	http://www.ebi.ac.uk/arrayexpress
BIOGRID	http://thebiogrid.org
Data Repository of Yeast Genetic Interactions	http://drygin.ccbr.utoronto.ca
European <i>Saccharomyces cerevisiae</i> Archive for Functional Analysis	http://web.uni-frankfurt.de/fb15/mikro/euroscarf/col_index.html
g:Profiler	http://biit.cs.ut.ee/gprofiler
GeneMANIA	http://www.genemania.org
Munich Information Center for Protein Sequencing	http://mips.helmholtz-muenchen.de/genre/proj/yeast
Protein-Protein Interaction Prediction Engeen	http://cgmlab.carleton.ca/PIPE
<i>Saccharomyces cerevisiae</i> Morphology Database	http://yeast.gi.k.u-tokyo.ac.jp/datamine
<i>Saccharomyces</i> Genome Database	http://www.yeastgenome.org
Stanford Microarray Database	http://smd.princeton.edu
The Gene Ontology	http://www.geneontology.org
Yeast Deletion Project Database	http://www-sequence.stanford.edu/group/yeast_deletion_project/
Yeast feature	http://software.dumontierlab.com/yeastfeatures
Yeast GFP Fusion Localization Database	http://yeastgfp.yeastgenome.org
Yeast Protein Localization Database	http://ypl.tugraz.at/pages/home.html
Yeast Resource Center	http://depts.washington.edu/yeastrc

Appendix 2-1: List of designed primes for gene knock-out in yeast

Primer name	Primer sequence (5'-3')	description
BSC2-F	ATTGTCTTTTTTCTTCCTTCCTTGCCAGGCATCTGCATCC ATTATTGCCTTTAGTTACACACATACGATTTAGGTGACAC	Upstream homology attached to forward NAT primer, gene knock out
BSC2-R	TTCTGTAGCAATAAGTGTTCCTTAGCAGTGCGTTGCGTTG CCTTGACCTTGATTATCAATAATACGACTCACTATAGGGA G	Downstream homology attached to reverse NAT primer, gene knock out
BSC2-Con	AAGCCAGGAAGTTAGCGCCGGG	Confirmation analysis for gene replacement
OLA1-F	TAGCTTTTTGTGTTACTTCTAGTATCCGTAAAAACAATCA AATAAATAGCACTGTAAAACACATACGATTTAGGTGAC AC	Upstream homology attached to forward NAT primer, gene knock out
OLA1-R	ATGATGATATATAGTAGGGAATAAACCAATAAAGAAA AGTCAACTTTAGTGACGAAAAATACGACTCACTATAGG GAG	Downstream homology attached to reverse NAT primer, gene knock out
OLA1-Con	GGCAGAATGCAGTTGTTGCGTG	Confirmation analysis for gene replacement
YNL040W-F	GTGACGCATATCGACGAAAGCACACAACAGCGTCAAAA ATTGATTAAA AGGTAAGTTATCCACATACGATTTAGGTGACAC	Upstream homology attached to forward NAT primer, gene knock out
YNL040W-R	TCACGATAAATACCTACCATTATGTATACAACAAATCTGC GTATACAAATGGCACATTCAATACGACTCACTATAGGG AG	Downstream homology attached to reverse NAT primer, gene knock out
YNL040W-Con	AGGGTAGTAACGTACGTGGACGAA	Confirmation analysis for gene replacement
Nat-Con	TCCAGTGCCTCGATGGCCTCGGCG	Confirmation analysis for gene replacement
PGK1-F	CAGACCATTCTTGCCATCT	housekeeping gene, qRT-PCR analysis
PGK1-R	CGAAGATGGAGTCACCGATT	housekeeping gene, qRT-PCR analysis
LacZ-F	TTGAAAATGGTCTGCTGCTG	Reporter gene, qRT-PCR analysis
LacZ-R	TATTGGCTTCATCCACCACA	Reporter gene, qRT-PCR analysis
Ploy T	TTTTTTTTTTTTTTTTTTT	Template for Reverse transcriptase

Appendix 2-2: Alphabetically ordered list of genes identified through large-scale X-gal based *β-galactosidase* assay for stop codon read-through. They are randomly reconfirmed through quantitative ONPG passed *β-galactosidase* assay. Sensitivity analysis for translation inhibitory drugs is also included.

Systematic Name	Standard name	Ratio of the β -gal for UAA stop codon bypass normalized to a control strain(wt) set at 1.00 (pUKC817/pUKC815)	Ratio of the β -gal for UGA stop codon bypass normalized to a control strain(wt) set at 1.00 (pUKC818/pUKC815)	Ratio of the β -gal for UAG stop codon bypass normalized to a control strain(wt) set at 1.00 (pUKC819/pUKC815)	Sensitivity to Cycloheximide	Sensitivity to Paromomycin
YJR105W	ADO1	9 ± 1.09	11 ± 2.11	5 ± 0.12	Yes	No
YCL025C	AGP1	-	-	-	yes	No
YDR063W	AIM7	-	-	-	No	No
YPR049C	ATG11	-	-	-	No	Yes
YJR148W	BAT2	-	-	-	No	No
YER167W	BCK2	-	-	-	Yes	No
YDR099W	BMH2	-	-	-	Yes	No
YDR275W	BSC2	-	-	-	Yes	Yes
YNR069C	BSC5	-	-	-	Yes	Yes
YKL011C	CCE1	9 ± 6.13	8 ± 10.46	9 ± 3.70	yes	No
YML071C	COG8	-	-	-	Yes	No
YIL036W	CST6	-	-	-	No	No
YHR028C	DAP2	-	-	-	No	Yes
YFR044C	DUG1	-	-	-	No	Yes
YHR021W-A	ECM12	-	-	-	No	No
YJR137C	ECM17	-	-	-	Yes	No
YMR031C	EIS1	-	-	-	Yes	No
YLR214W	FRE1	-	-	-	Yes	No
YOR381W	FRE3	-	-	-	Yes	No
YOR324C	FRT1	1 ± 0.08	2 ± 1.01	1 ± 0.43	Yes	No
YOL049W	GSH2	-	-	-	yes	No
YLL060C	GTT2	-	-	-	No	No
YJR069C	HAM1	-	-	-	Yes	No

YNL281 W	HCH1	-	-	-	Yes	No
YLR192C	HCR1	-	-	-	Yes	Yes
YLL019C	KNS1	-	-	-	Yes	No
YDR503C	LPP1	-	-	-	Yes	No
YGL197 W	MDS3	-	-	-	Yes	No
YFR030W	MET10	-	-	-	No	Yes
YNL145 W	MFA2	10 ± 1.49	12 ± 9.58	2 ± 0.08	No	No
YER028C	MIG3	-	-	-	Yes	No
YIR025W	MND2	-	-	-	No	No
YLR219W	MSC3	-	-	-	Yes	No
YMR080 C	NAM7	-	-	-	Yes	No
YBR212 W	NGR1	-	-	-	No	No
YDR456 W	NHX1	-	-	-	Yes	No
YHR077C	NMD2	-	-	-	Yes	No
YOR310C	NOP58	-	-	-	No	Yes
YJR062C	NTA1	-	-	-	No	Yes
YAL015C	NTG1	7 ± 4.11	11 ± 8.21	2 ± 0.87	Yes	No
YBR025C	OLA1	-	-	-	Yes	Yes
YIL136W	OM45	-	-	-	No	No
YHL020C	OPI1	-	-	-	Yes	No
YLR037C	PAU23	-	-	-	No	Yes
YMR231 W	PEP5	-	-	-	No	Yes
YIL037C	PRM2	-	-	-	No	No
YHR189 W	PTH1	-	-	-	Yes	Yes
YMR274 C	RCE1	-	-	-	Yes	No
YOR275C	RIM20	-	-	-	Yes	No
YDR198C	RKM2	9 ± 2.93	5 ± 0.52	2 ± 0.85	Yes	No
YPL220W	RPL1A	-	-	-	No	Yes
YDL075 W	RPL31A	-	-	-	Yes	No
YDL076C	RXT3	-	-	-	No	No
YAL027 W	SAW1	-	-	-	No	Yes
YER087C -B	SBH1	-	-	-	No	No
YKL148C	SDH1	-	-	-	Yes	No
YLL041C	SDH2	-	-	-	No	No
YHR207C	SET5	-	-	-	Yes	No
YDR477 W	SNF1	-	-	-	No	No

YMR183 C	SSO2	12 ± 2.54	8 ± 1.47	3 ± 0.68	Yes	No
YPR095C	SYT1	-	-	-	Yes	No
YNL128 W	TEP1	-	-	-	No	No
YPR040W	TIP41	-	-	-	Yes	No
YER007C -A	TMA20	-	-	-	No	Yes
YJR014w	TMA22	-	-	-	No	No
YER113C	TMN3	-	-	-	No	No
YOL006C	TOP1	-	-	-	Yes	No
YGR072 W	UPF3	-	-	-	Yes	No
YPR152C	URN1	-	-	-	No	No
YIL056W	VHR1	4 ± 1.06	10 ± 2.57	3 ± 0.52	No	No
YOR230 W	WTM1	-	-	-	Yes	No
YLL048C	YBT1	-	-	-	Yes	No
YER156C	YER156C	-	-	-	Yes	No
YFL043C	YFL043C	-	-	-	Yes	No
YGL072C	YGL072C	-	-	-	Yes	Yes
YGR117C	YGR117C	-	-	-	No	No
YIL032C	YIL032C	-	-	-	Yes	No
YLR445W	YLR445W	-	-	-	No	Yes
YMR031 W-A	YMR031 W-A	-	-	-	No	No
YOR064C	YNG1	1 ± 0.03	2 ± 1.28	1 ± 0.07	Yes	Yes
YNL040 W	YNL040 W	-	-	-	Yes	Yes
YNL122C	YNL122C	-	-	-	Yes	No
YOR378 W	YOR378 W	-	-	-	Yes	No
YLR120C	YPS1	-	-	-	No	No

Appendix 2-3: SGA results of synthetic sick interactions for Ola1, Bsc2 and YNL040W.

* Represents interactions included from literatures.

SGA-<i>ola1</i>Δ				SGA-<i>bsc2</i>Δ		SGA-<i>ynl040w</i>Δ		
ACT1*	ERG25*	MUD2	RPS10A	ARC1		AIM22*	MCX1*	TMA17*
ALG9*	ERI1*	NSR1	RPS10B	ASM4*		ALT2*	MGR3*	UPS3
ALT1	GCN1	NUT1	RRN10	BIO2		APT1*	NAP1*	VMA8*
ARC1*	GCN4	PBS2*	SAC3	CET1*		ATG17*	NAT4*	VPS5*
ARC18*	GIM5*	PMP3*	SCP160	CHO2*		CBS2	NOP13	WBP1*
ARD1	GPX2*	PTK1*	SGN1	EAF7		CDC2	NOP16	XRS2*
ATG11*	GYP6*	PUF2	SHE4	HMG1		CSR1*	PCI8*	YIF1*
BAT2	HDA3*	PUF3	SIG1	PBS2*		CUP5*	RPL20B	
BSC4	HOT1	RGD1*	SIT4	PET54		EAP1	RPL31A	
BUD27*	IKI3	RML2	SKI3	RAD27*		EDC2	RPS30A	
BZZ1	ILM1	RPL12A	SKI8	SCS7*		EMI2*	RRM3*	
CCR4	ILV1	RPL12B	SPC24*	SLM1		FRK1	RTS1	
CDC37*	ILV6	RPL15B	SRF1*	SWF1*		GDE1	SEM1	
CLU1	INO80*	RPL20B	TPA1	THG1		HAL5*	SGN1	
CYS4	JSN1	RPL22B	TRS33	THG18*		HIR3*	SSD1	
DEP1*	KGD2*	RPL27A	UPF3	TUM1		HUR1*	SUL1	
DGK1*	KRS1*	RPL27B	VID22	UPF3		IES2*	SUT2*	
DHH1	LEA1*	RPL31A	VTC1	UTR1*		IRS4*	SVL3*	
DUO1*	LOC1	RPL33B	YNL171C	YSC83*		KGD2*	TFP1*	
EAF7	LSM6	RPL34B	YOR302W					
EAP1	MDM34	RPL37A	ZUO1					
EAR1	MET14	RPL37A						
EFT1	MET17	RPL40B						

Appendix 2-4: Re-introduction of deleted target genes in double mutants. A) Re-introduction of target genes reverses the observed genetic interactions phenotype. The sick phenotypes of double gene knockout strains are reversed when target genes are re-introduced into the corresponding mutant strains. Compensation for double gene knockout mutant strains, *bsc5Δ-rps17aΔ* and *bsc5Δ-ccr4Δ*, are used as negative controls. B) A representative figure illustrating the reversion of sick phenotypes by re-introduction of corresponding genes.

A

Double mutant strains	Phenotype		
<i>ola1Δ-dhh1Δ</i> <i>ola1Δ-rpl15bΔ</i> <i>ola1Δ-puf2Δ</i>		p-OLA1	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
	Sick	WT	Sick
<i>ynl040wΔ-nop16Δ</i> <i>ynl040wΔ-rpl20bΔ</i> <i>ynl040wΔ-ssd1Δ</i>		p-YNL040W	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
	Sick	WT	Sick
<i>bsc2Δ-pet54Δ</i> <i>bsc2Δ-tum1Δ</i> <i>bsc2Δ-eaf7Δ</i>		p-BSC2	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
	Sick	WT	Sick

Control double mutant strains	Phenotype		
<i>bsc5Δ-rps17aΔ</i> <i>bsc5Δ-ccr4Δ</i>		p-OLA1	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-rps17aΔ</i> <i>bsc5Δ-ccr4Δ</i>		p-YNL040W	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-rps17aΔ</i> <i>bsc5Δ-ccr4Δ</i>		p-BSC2	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick

Appendix 3-1: List of primers used in this study.

Primer name	Sequence	
BS-MRPS16-F	AAGTAACGGAGGAGTTCTGACTGGTATTCGGTTAAGAAATTG TAATAGAGTTTATCAAGTCACATACGATTTAGGTGACAC	Gene knock-out
BS-MRPS16-R	ATTAACACACGTTCTTTGCTTCTTTCTTTCTTTCTTTCTTTT TTTCCGATCTTTAATACGACTCACTATAGGGAG	Gene knock-out
BS-MRPS16-con	GAGGGGCATTAGCCTTTCTGCAG	Gene knock-out
BS-VPS70-F	GCAAGGGGCCGTACTTATTAACAATTTCAACATGCACAAT TAAGATGTAAGAGATATCACATACGATTTAGGTGACAC	Gene knock-out
BS-VPS70-R	AAGTCACTCTACGTATGCAAATAAATAAATGTATATACTGA GGTGAGGTACTTAAATGCAATACGACTCACTATAGGGAG	Gene knock-out
BS-VPS70-con	GACAAGGCAAGGGGCCGTACTTAT	Gene knock-out
BS-YBR064W-F	TTTCTCTTCGGAACGTCACTTCGATTATTGAGGCAGTCTTC ATTAAAGTTTATTACACATACGATTTAGGTGACAC	Gene knock-out
BS-YBR064W-R	AGGTACATGATATATTGACTGATTTCTGATAGCTCTTTTTTTC CAATATCGATAAAAAAATACGACTCACTATAGGGAG	Gene knock-out
BS-YBR064W-con	CGGGTAGGAGAGTTGGTCACCAAA	Gene knock-out
BS-RPL36A-F	CAGTATAGTACTGATTTACAAGTATCAAAGAAACAAGAAAGA CAGCAAACACAATAAGAACACATACGATTTAGGTGACAC	Gene knock-out
BS-RPL36A-R	CTAAGTTAAAAAGAAAATATGATATCTATATATATATTACAAA AATCAGTTTTTAGAAACAATACGACTCACTATAGGGAG	Gene knock-out
BS-RPL36A-con	GAGAATCCCCACATAAGGATGCA	Gene knock-out
BS-PCS60-F	TTGAAATTTACTCCCACTGTGCTGCAAGAGGTTAGACGCTTCA CGCACATTTTGTACACACATACGATTTAGGTGACAC	Gene knock-out
BS-PCS60-R	GTATATAAACGGGTGCCAGATAGAAGCTTTCAGAGAGCATAA AATTGTACAGGATACTGCAATACGACTCACTATAGGGAG	Gene knock-out
BS-PCS60-con	CCAAACAGCATGGGGCACCGTTTC	Gene knock-out
BS-ATP22-F	CGTACTGGCAAGAATGGGCTTTCAAATCTTCCAAGCGTTCAT TTAATGTTCTTATTGGCCACATACGATTTAGGTGACAC	Gene knock-out
BS-ATP22-R	TTATGGGTAAATTACTATCTATGAATGAATATTTACTATTTACT AGTGCTCATCTGGATAAATACGACTCACTATAGGGAG	Gene knock-out
BS-ATP22-Con	CTAGGTTGGAGCTAGGGCGTCTTG	Gene knock-out
Nat-Con	TCCAGTGCCTCGATGGCCTCGGCG	Gene knock-out
HSC82-F	CCGGTGAATCTCTAAAGGCA	qRT-PCR
HSC82-R	ATCAATTGGGTCCGGTCAAGA	qRT-PCR
HSP82-F	GAGTTGACGAAGGTGGTGCT	qRT-PCR
HSP82-R	ATGCAAAGGAAGTTGGTTCG	qRT-PCR
PGK1-F	CAGACCATTCTTGCCATCT	qRT-PCR (house keeping)
PGK1-R	CGAAGATGGAGTCACCGATT	qRT-PCR (house keeping)
LacZ-F	TTGAAAATGGTCTGCTGCTG	qRT-PCR
LacZ-R	TATTGGCTTCATCCACCACA	qRT-PCR
Ploy T	TTTTTTTTTTTTTTTT	CDNA preparation

Appendix 3-2: The combination (from this study and the literature) of synthetic sick interaction for *ATP22*, *MRPS16*, *PCS60*, *RPL36A*, *YBR063C* and *DOM34*. Genetic interaction data is quantified and color coded. ^{*} Represents interactions that were included from literature. ^A Represents conditional SGA under acetic acid treatment (180mM). ^C Represents conditional SGA under cycloheximide (40 ng/ml) treatment. ^H Represents conditional SGA under heat shock (37 °C) condition. ^P Represents conditional SGA under paromomycin (18 mg/ml) condition. ^R Represents conditional SGA under rapamycin condition (4 ng/ml). ^S Represents SGA data with no condition.

ATP22

MRPS16

AE2 ^S		RPS4B ^P	
ANB1 ^{P,S}		RPS8A ^P	
ATP22 ^A		RPS9B ^P	
BRR1 ^R		RSA1 ^P	
BUD21 ^P		SAC3 ^A	
BUD27 ^A		SCD6 ^A	
CDH1 ^{P,S}		STF1 ^A	
CIS3 ^R		SWS2 ^S	
CLN3 ^A		TEF4 ^A	
DST1 ^R		TIF1 ^R	
EAP1 ^P		TMA20 ^{A,P}	
ECM1 ^R		TOS6 ^A	
EDC2 ^A		UBP3 ^P	
EFM2 ^A		UME6 ^A	
GLC3 ^{C,P}		VPS35 ^R	
GLC3 ^P		WHI3 ^{A,S}	
GUF1 ^R		WHI5 ^{A,P,R}	
HHT1 ^{P,R,A}		YAK1 ^A	
HPM1 ^S		YAL012W ^A	
NTH1 ^S		YAR1 ^A	
RBG2 ^S		YBR064W ^A	
RKM2 ^{P,S}		YBR089C-A ^S	
RPL12B ^A		YDR382W ^{A,C,P}	
RPL16A ^R		YGL035C ^A	
RPL24A ^P		YGR285C ^P	
RPL34A ^R		YJL080C ^S	
RPL37B ^P		YJL089W ^{A,C,S}	
RPL41A ^R		YJL160C ^R	
RPL42A ^R		YJL176C ^{P,R}	
RPL6B ^A		YKR006C ^P	
RPL8B ^P		YLR085C ^S	
RPS10A ^{A,R}		YOL115W ^P	
RPS11B ^A			
RPS17A ^P			
RPS23B ^R			
RPS24B ^P			
RPS26B ^R			
RPS27B ^P

RPL36A

AAP1 ^S		RIC1 ^P		SAC3 ^P		YGR053C ^P	
APE2 [*]		RMD1 ^A		SAF1 ^P		YGR285C ^{A, C, H, P, R}	
ARD1 [*]		RML2 ^P		SEC17 [*]		YHR121W ^A	
ARX1 ^P		RPA14 ^A		SHF1 [*]		YJL051W ^A	
ASC1 ^P		RPL12B ^P		SIL1 ^A		YJL056C ^P	
BUD23 ^P		RPL13A ^A		SIP2 ^P		YJL080C ^A	
CDH1 ^S		RPL16B ^P		SLX9 ^P		YJL089W ^R	
CLN3 ^P		RPL19A ^{C, P}		SSB1 ^P		YJL095W ^R	
COG3-1 [*]		RPL20B ^S		SSF2 ^R		YJL124C ^A	
COX23 [*]		RPL22A ^P		STB3 ^A		YJL176C ^A	
DOA1 [*]		RPL22B ^P		STF1 ^A		YKL205W ^A	
DPH2 ^P		RPL24A ^A		TAE2 ^R		YLR050C [*]	
DST1 ^{P, S}		RPL27A ^P		THP2 [*]		YLR451W ^P	
EAP1 ^{A, C, P, R}		RPL27B ^A		TMA19 [*]		YMR125W ^P	
ECM1 ^P		RPL29 ^P		TMA64 ^P		YMR172W ^P	
EDC2 ^P		RPL2A ^P		TOM1 [*]		YMR190C ^P	
ERG11 [*]		RPL34A ^S		TRM1 ^P		YMR194C-A ^P	
GLC3 ^P		RPL35B ^A		UBC4 [*]		YMR31 ^A	
GSY1 ^{A, R}		RPL37B ^A		UBP3 ^{P, *}		YOR115C ^P	
HHT1 ^{A, C, P}		RPL43A ^P		UTR2 [*]		YOR302W ^P	
HPM1 ^P		RPL6B ^{A, H, P, R}		VPS27 [*]		YPL086C ^R	
HTZ1 ^P		RPN11 [*]		VPS29 [*]		YPL183W-A ^P	
INO1 ^R		RPP1A ^P		WHI3 ^{A, P, R}			
ISW2 [*]		RPS0A ^P		WHI5 ^{A, S}			
JNM1 [*]		RPS0B ^A		XRN1 ^P			
LOC1 ^P		RPS10A ^A		YAL012W ^P			
LTV1 ^P		RPS16A ^P		YAL029C ^A			
MRP49 ^P		RPS17A ^A		YBR057C ^A			
MRPL39 ^P		RPS18A ^P		YDL116W ^P			
NTH1 ^{A, P}		RPS19B ^A		YDR282C ^A			
NUP2 ^P		RPS1A ^A		YDR382W ^P			
PAT1 [*]		RPS1B ^P		YDR515W ^P			
PCH2 ^P		RPS27B ^A		YEL020C [*]			
PCI8 [*]		RPS29B ^S		YER086W ^P			
PRC1 ^{A, S}		RPS30B ^P		YER088C ^P			
PXA2 [*]		RPS4B ^A		YFL031W ^P			
RAD61 [*]		RPT2 [*]		YGL013C ^P			
REI1 ^P		RSA1 ^P		YGL035C ^P			

YBR063C

DOM34

AAH1 ^C		PCI8 ^{A, C}		RPS24B ^{*, A}		TRP1 ^A		YNL269W ^{A, R}	
AIM23 ^A		PET54 ^C		RPS26B ^A		TSA1 ^{C, R}		YNL277W ^P	
ARX1 ^{A, P}		PHO80 ^R		RPS27B ^P		UBP3 ^P		YOR017W ^C	
BRR1 ^{A, R}		PRC1 ^A		RPS28A [*]		UME6 ^P		YOR028C ^A	
BUD23 ^P		RBG2 ^A		RPS28B [*]		VMA3 [*]		YOR302W ^{P, R}	
CDH1 ^{A, S}		REI1 ^P		RPS29A [*]		VPS35 ^{C, P}		YPL048W ^P	
CYC1 ^C		RMD1 ^A		RPS29B ^{*, P}		XRN1 ^{A, P}		ZDS1 ^C	
DST1 ^P		RPA14 ^{A, P}		RPS30A [*]		YAL012W ^P			
EFT2 ^A		RPA34 ^{A, S}		RPS30B [*]		YAP1 ^{R, *, A, C, H}			
GCN3 ^A		RPL12B ^P		RPS4A ^A		YBL079W ^{A, C}			
GIS2 ^C		RPL13A ^P		RPS4B ^{*, P}		YBR062C ^{A, P, R}			
GLC3 ^A		RPL16B ^P		RPS6A [*]		YCL009C ^P			
GLG2 ^R		RPL21B ^P		RPS6B [*]		YCL029C ^A			
GRX3 ^A		RPL22A ^A		RPS7A ^P		YDL037C ^A			
GSY1 ^P		RPL23A ^{A, C, P}		RPS7B ^C		YDL130W ^A			
GUF1 ^R		RPL24A ^{A, C, P}		RPS8A ^{A, P}		YDR382W ^P			
HCR1 ^P		RPL24B ^C		RPS9B ^P		YDR515W ^P			
HHF1 ^{A, C, P}		RPL34B ^P		RSA1 ^P		YDR535C ^{A, R}			
HHO1 ^A		RPL36A ^C		RSM25 ^A		YEL048C ^C			
HPM1 ^P		RPL37B ^{A, C}		RSM27 ^A		YER086W ^P			
HSC82 ^R		RPL9A ^{A, P}		SAC3 ^P		YJL051W ^P			
HSP30 ^A		RPP1A ^P		SBP1 ^A		YJL056C ^{A, C}			
HTA1 ^R		RPS0A ^P		SDA1 [*]		YJL124C ^A			
IMG2 ^{A, P}		RPS0B ^C		SER33 ^C		YJL152W ^C			
INO1 ^C		RPS10A ^A		SHE4 ^R		YJL176C ^A			
LOC1 ^P		RPS10A ^{R, S, H}		SLX9 ^{A, P}		YJR003C [*]			
LTV1 ^P		RPS11B ^A		STF1 ^A		YJR137C ^{A, P}			
MCK1 ^{C, R}		RPS14A ^P		STM1 ^R		YKL074C ^P			
MRPL1 ^A		RPS16A ^{A, P}		SYS1 [*]		YKL205W ^A			
MRPL10 ^R		RPS17A ^P		TAE1 ^A		YLR107W ^A			
MRPL9 ^R		RPS18B ^{A, P}		TGS1 ^{A, C, S}		YLR434C [*]			
MRPS17 ^P		RPS19A ^P		THI2 ^A		YLR451W ^A			
NOP12 ^P		RPS1A ^{*, A, C, P, R}		TIF4632 ^C		YMR125W ^P			
NOP16 ^P		RPS1B ^{*, P}		TMA20 ^{*, P}		YMR171C ^{A, R}			
NTH1 ^{A, R}		RPS21A ^P		TMA7 ^R		YMR190C ^{A, C}			
OAZ1 ^C		RPS22B [*]		TOR1 ^{A, R}		YMR307W ^C			
PBP1 ^A		RPS23A [*]		TPK1 ^A		YMR31 ^{C, P}			
PCH2 ^P		RPS24A ^P		TRM1 ^{A, R}		YNL087W ^P			

Appendix 3-3: Re-introduction of deleted target genes in double mutants. Re-introduction of target genes reverses the observed genetic interactions phenotype. The sick phenotypes of double gene knockout strains are reversed when target genes are re-introduced into the corresponding mutant strains. Compensation for double gene knockout mutant strains, *bsc5Δ-ser2Δ* and *pop2Δ-cka2Δ*, are used as negative controls.

Double mutant strains	Phenotype		
<i>atp22Δ-rps17aΔ</i> <i>atp22Δ-tef4Δ</i>		p-ATP22	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
<i>vps70Δ-hht1Δ</i> <i>vps70Δ-rps10aΔ</i>		p-VPS70	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
<i>dom34Δ-rpl37bΔ</i> <i>dom34Δ-pci8Δ</i>		p-DOM34	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
<i>mrps16Δ-anb1Δ</i> <i>mrps16Δ-rkm2Δ</i>		p-MRPS16	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
<i>pcs60Δ-rsa1Δ</i> <i>pcs60Δ-rpl43aΔ</i>		p-PCS60	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
<i>rpl36aΔ-rpl20bΔ</i> <i>rpl36aΔ-aap1Δ</i>		p-RPL36A	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick
<i>ybr063cΔ-rpl23aΔ</i> <i>ybr063cΔ-efi2Δ</i>		p-YBR063C	Empty plasmid
	Sick	WT	Sick
	Sick	WT	Sick

Control double mutant strains	Phenotype		
<i>bsc5Δ-ser2Δ</i> <i>pop2Δ-cka2Δ</i>		p-ATP22	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-ser2Δ</i> <i>pop2Δ-cka2Δ</i>		p-VPS70	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-ser2Δ</i> <i>pop2Δ-cka2Δ</i>		p-DOM34	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-ser2Δ</i> <i>pop2Δ-cka2Δ</i>		p-MRPS16	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-ser2Δ</i> <i>pop2Δ-cka2Δ</i>		p-PCS60	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-ser2Δ</i> <i>pop2Δ-cka2Δ</i>		p-RPL36A	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick
<i>bsc5Δ-ser2Δ</i> <i>pop2Δ-cka2Δ</i>		p-YBR063C	Empty plasmid
	Sick	Sick	Sick
	Sick	Sick	Sick

Appendix 3-4: List of gene deletion mutants compensated by overexpression plasmids in different stress condition.

ATP22 overexpress		MRPS16 overexpress		
Cycloheximide	Heat shock	Cycloheximide	Rapamycin	Heat shock
ASC1	BRR1	ASC1	ANB1	BRR1
EAP1	BUD21	BUD23	BUD21	BUD21
MRP17	BUD27	DST1	CYC1	BUD27
NOP16	ISY1	NOP16	EAP1	IMG2
RPL22A	LOC1	RPL20B	GAS2	ISY1
TEF4	OAZ1	RPS0A	HAP4	LOC1
TRP1	PET54	RPS9B	HSP82	OAZ1
YDR535C	TMA7	STF1	ISY1	PET54
YER081W	YDR515W	TEF4	JJJ1	YDR515W
YMR172W	YNL087W	YDR535C	MRN1	YNL087W
YPL086C	Acetic Acid	YER081W	NEW1	YNR048W
Rapamycin	URE2	YJR148W	RKM4	paromomycin
DHH1	ATP22	YLR107W	RKM5	ASC1
EAP1	DPH5	YMR172W	RPL12B	GLC3
GAS2	GBP2	YPL086C	RPL17B	RPL20B
HAP4	HSC82	Acetic Acid	RPL27A	RPS18A
HSC82	HSP82	DTD1	RPL31B	RPS30B
LTV1	RPL22A	ECM1	RPL36A	RPS9B
NTH1	RPL27A	GAS2	RPL38	RSM27
RPL12B	RPL31B	HSC82	RPP1A	
RPL36A	RPS18A	HSP82	RPS18A	
RPL8B	RSA3	LOC1	RPS1A	
RPP1A	RSM27	MRPS16	RPS8A	
RPS18A	SHE4	PET123	SAC3	
RPS1A	TEF4	PRC1	STM1	
RPS7B	YIL096C	RPL29	TAE2	
RPS8A	YLR149C	RPS17B	TIF4632	
URN1	YOP1	RPS18A	URE2	
VPS70		RSA3	URN1	
YJL095W		TEF4	VPS35	
YJR148W		YFL031W	YFL034W	
YKL068W		YHR077C	YIL096C	
YKR020W		YIL096C	YKL068W	
YPL067C		YJL051W	YLR149C	
paromomycin		YJL089W	YPR014C	
ASC1		YMR179W		

RPS10A	YOR047C
RPS16A	
RPS18A	
RPS9B	

PCS60 overexpress		RPL36A overexpress	
Cycloheximide	Rapamycin	Cycloheximide	Rapamycin
ASC1	ANB1	ASC1	YIL096C
BUD23	CYC1	DPH2	ANB1
DST1	EAP1	NOP16	BUD21
NOP16	GAS2	RMT2	CYC1
RPL20B	HAP4	RPS29B	HAP4
RPS17A	HSC82	TEF4	HSC82
TEF4	HSP82	YDR535C	HSP82
YDR535C	JJJ1	YER081W	ISY1
YER081W	MRP20	YKL205W	JJJ1
YJR148W	MRPL9	YMR172W	MRN1
YMR172W	NEW1	YPL086C	NEW1
YPL086C	PET123	Heat shock	RKM5
Acetic Acid	RKM5	BRR1	RPL17B
BUD21	RPL12B	BUD21	RPL24B
CDC26	RPL12B	BUD27	RPL27A
GAS2	RPL17B	IMG2	RPL31B
HSC82	RPL27A	ISY1	RPL38
HSP82	RPL31B	OAZ1	RPP1A
LOC1	RPP1A	PET54	RPS18A
MRPL38	RPP1A	YDR515W	RPS1A
PCS60	RPS8A	YNL087W	RPS7B
RPB9	SAC3	Acetic Acid	STM1
RPP1A	TAE2	ALB1	TAE2
RPS18A	TIF4632	ANB1	TIF4632
RPS23B	URE2	BUD23	TOR1
RPS27B	YFL034W	GAS2	TRP1
RSA3	YGR285C	HSC82	URE2
TEF4	YJL095W	HSP82	URN1
YAR1	YPL183W-A	LOC1	VPS35
YDL012C	YPR014C	PET123	VPS70
YDL088C	paromomycin	RPL36A	YFL034W

YIL096C	ASC1	RPL36A	YKL068W
YJL051W	RPL20B	RPP1A	YLR149C
YPL048W	RPS10A	RPS18A	YOL114C
Heat shock	RPS17A	TEF4	YPL086C
BRR1	RPS18A	YAR1	YPR014C
BUD21	RPS4B	YDL088C	paromomycin
BUD27	RPS9B	YFL031W	ASC1
CDC26	YGR285C	YIL096C	HSP82
IMG2		YJL152W	RPS18A
LOC1			RPS9B
SHE4			YHR121W
YDR515W			
YNL087W			

VPS70 overexpress		YBR063C overexpress		DOM34 overexpress	
Cycloheximide	Acetic Acid	Cycloheximide	Acetic Acid	Cycloheximide	Acetic Acid
ASC1	GAS2	ASC1	GAS2	ASC1	DOM34
EAP1	HSC82	RMD1	HSC82	RPL21B	GAS2
NOP16	HSP82	RPL6B	HSP82	YER081W	HSC82
RMD1	LOC1	RPS17A	PET123	YMR172W	HSP30
RPS9B	MRP49	TEF4	PRC1	YOL115W	HSP82
TEF4	PRC1	TMA20	RPL27A	YPL086C	LOC1
YER081W	RPL34A	YDR382W	RPS18A	Rapamycin	NCL1
YGR285C	RPS18A	YER081W	RSA3	ANB1	PET123
YMR172W	RPS9B	Rapamycin	RSM27	BRR1	RPS21A
Rapamycin	RSA3	ANB1	SSF2	CYC1	RSA3
GLG1	SAC3	BUD21	TEF4	GBP2	SUC2
HAP4	TEF4	CYC1	YBR063C	HAP4	YAR018C
HSC82	VPS70	DBP7	YBR089C-A	HHO1	YDL088C
JJJ1	YIL096C	HAP4	YIL096C	JJJ1	YDR378C
RKM5	YJL051W	HSC82	YJL051W	MET17	YIL096C
RPL24B	YJR137C	HSP82	YJL152W	NEW1	YIL096C
RPS1A	paromomycin	ISY1	YMR129W	RKM4	YPL048W
RPS7B	ASC1	JJJ1	Heat shock	RPL12B	paromomycin
TAE2	RPL35B	MRN1	BRR1	RPL17B	ASC1
URE2	RPS10A	NEW1	BUD21	RPL27A	BUD23
YIL096C	RPS16A	RKM5	ISY1	RPL37B	RPL43A
Heat shock	RPS18A	RPL12B	OAZ1	RPS1A	RPS16A
ISY1	RPS9B	RPL17B	RPL36A	TAE2	RPS9B
LOC1	YGR285C	RPL24A	YDR515W	TIF4632	YER081W

OAZ1 YDR515W YNL087W YPR197C	RPL24B	YNL087W	URE2
	RPL38	paromomycin	URN1
	RPL6B	DPH5	YAK1
	RPP1A	MRPL40	YLR149C
	RPS18A	RPL41A	YNL323W
	RPS1A	SSF2	YPL067C
	STM1	YLR085C	YPL197C
	TAE2		YPR014C
	TIF4632		YPR197C
	URE2		Heat shock
	URN1		BRR1
	YFL034W		BUD21
	YIL096C		CDC26
	YKL068W		IMG2
	YKR020W		ISY1
	YLR149C		LOC1
	YPR014C		OAZ1
			PET54
			YAL012W

Appendix 4-1: Precision-Recall curve for *H. sapiens* PPIs. The curves presents the Recall (True Positive Rate) against Precision including homologs (dashed blue line) and with homologs removed (dotted pink line).

Appendix 4-2: MP-PIPE benchmark performance. A) Distribution of running times for human protein-protein interaction prediction. Numbers above bars indicate approximate number of protein pairs with a running time within the given range. B) Performance for different numbers of threads per worker on the large cluster. Average running times for 5,000 random protein pairs (small number of threads) or 50,000 random protein pairs (large number of threads), using one worker process. C) Performance for different numbers of workers on the large cluster. Average running times for 500,000 random protein pairs, using 512 threads per worker.

Appendix 4-3: List of *Homo sapiens* interactions predicted in this study. Each prediction is accompanied by a score, if the interaction was previously reported (known) or novel, if both proteins are in the same component, have the same function or involved in the same process (GO ontology) as well as if they share a third party interaction.

Protein A	Protein B	Score	Novel / Known	Same component	Same function	Same process	Third Party Interaction
Q5VST9	Q8WZ42	0.99999999	Known	YES	YES	YES	YES
P20929	Q8WZ42	0.99999998	Known	YES	YES	NO	YES
P21333	Q8WZ42	0.99999996	Known	YES	YES	NO	NO
P35579	Q8WZ42	0.99999994	Known	YES	YES	NO	YES
P16157	Q8WZ42	0.99999994	Known	YES	NO	NO	YES
P52179	Q8WZ42	0.99999993	Known	NO	YES	NO	YES
P54296	Q8WZ42	0.99999992	Known	NO	YES	NO	YES
Q86TC9	Q8WZ42	0.99999991	Known	YES	NO	NO	YES
Q14896	Q8WZ42	0.99999991	Known	NO	YES	NO	YES
Q00872	Q8WZ42	0.9999999	Known	NO	YES	NO	NO
Q14596	Q8WZ42	0.99999988	Known	NO	NO	NO	NO
Q01484	Q5VST9	0.99999987	Known	YES	YES	NO	NO
P35609	Q8WZ42	0.99999987	Known	YES	YES	NO	YES
P12814	Q8WZ42	0.99999987	Known	YES	YES	NO	NO
P20807	Q8WZ42	0.99999985	Known	YES	YES	NO	NO
P21817	Q92736	0.99999984	Known	YES	YES	YES	YES
Q15413	Q92736	0.99999983	Known	YES	YES	YES	NO
P08519	P98164	0.99999981	Known	NO	NO	YES	YES
P04114	P98164	0.99999981	Known	YES	NO	NO	NO
O60494	P98164	0.99999976	Known	YES	YES	YES	YES
O94833	Q9NYQ7	0.99999976	Known	NO	YES	NO	NO
Q5VST9	Q9Y623	0.99999974	Novel	YES	YES	NO	NO
Q5VST9	Q9UKX2	0.99999974	Known	YES	YES	NO	NO
P16157	Q5VST9	0.99999973	Known	YES	NO	NO	YES
Q14643	Q92736	0.9999997	Known	YES	YES	YES	YES
P25391	P98160	0.9999997	Known	YES	NO	NO	YES
Q13439	Q96PK2	0.99999969	Known	YES	NO	NO	NO
P01266	P98164	0.99999969	Known	NO	NO	NO	NO
O95613	Q99996	0.99999969	Known	YES	NO	NO	YES

P60709	Q8WZ42	0.99999967	Novel	YES	YES	NO	NO
P68133	Q8WZ42	0.99999967	Known	YES	YES	YES	NO
Q8WZ42	Q9GZV1	0.99999966	Known	NO	YES	NO	YES
Q09472	Q15751	0.99999966	Known	YES	NO	NO	NO
P98164	Q9Y490	0.99999966	Known	NO	YES	NO	NO
Q8WZ42	Q969Q1	0.99999965	Known	YES	NO	NO	NO
Q96PK2	Q9Y4D7	0.99999965	Known	NO	NO	NO	NO
Q13813	Q15149	0.99999965	Known	YES	YES	NO	YES
Q13023	Q92736	0.99999965	Known	YES	NO	NO	YES
Q12955	Q9H254	0.99999964	Known	YES	YES	NO	NO
P02549	Q15149	0.99999964	Known	YES	NO	NO	NO
Q13936	Q92736	0.99999963	Known	YES	YES	YES	YES
P98161	Q9Y490	0.99999963	Known	NO	YES	NO	YES
P02751	P08519	0.99999963	Known	YES	NO	NO	NO
O15230	Q02388	0.99999963	Known	YES	NO	NO	NO
P00451	Q07954	0.99999962	Known	YES	YES	NO	YES
Q15327	Q8WZ42	0.99999961	Known	YES	NO	NO	YES
P51587	Q13315	0.99999961	Known	YES	NO	YES	YES
P02751	P98161	0.99999961	Known	NO	NO	NO	YES
P02751	P98160	0.99999961	Known	YES	NO	YES	YES
Q96PK2	Q9P2S2	0.9999996	Known	NO	YES	NO	NO
Q14686	Q8NEZ4	0.9999996	Known	YES	YES	NO	YES
Q86SG2	Q8WZ42	0.99999959	Known	YES	NO	NO	YES
Q03164	Q92793	0.99999958	Known	YES	YES	YES	YES
P13611	P35555	0.99999958	Known	YES	YES	NO	YES
P21817	Q13698	0.99999957	Known	YES	YES	YES	NO
O94833	P16144	0.99999957	Known	NO	YES	YES	YES
Q96T58	Q9Y618	0.99999956	Known	YES	YES	NO	YES
Q01082	Q01484	0.99999956	Known	YES	NO	NO	NO
P98164	Q9NY46	0.99999956	Known	YES	NO	NO	NO
Q07954	Q9NY46	0.99999955	Known	NO	NO	NO	NO
Q14192	Q8WZ42	0.99999955	Known	YES	YES	NO	NO
P51587	Q13535	0.99999955	Known	YES	YES	YES	YES
P25391	Q02388	0.99999955	Known	YES	NO	NO	NO
P21333	P51587	0.99999955	Known	YES	NO	NO	YES
O75376	Q96T58	0.99999955	Known	NO	YES	YES	YES
Q96SN8	Q9P2P6	0.99999954	Known	NO	NO	NO	NO
Q96PK2	Q9ULB1	0.99999954	Known	NO	YES	NO	NO
Q12955	Q99250	0.99999954	Known	NO	NO	NO	NO
P20929	Q86TC9	0.99999954	Known	YES	YES	NO	YES
O75445	Q01955	0.99999953	Known	YES	NO	YES	NO

P02511	Q8WZ42	0.99999925	Known	YES	NO	NO	NO
P14543	P98160	0.99999925	Known	YES	YES	YES	YES
P07996	P98164	0.99999925	Known	NO	YES	NO	NO
Q13813	Q9Y566	0.99999924	Known	YES	NO	NO	YES
Q6UYC3	Q8NF91	0.99999923	Novel	NO	NO	NO	NO
P23229	Q15149	0.99999923	Known	YES	NO	NO	YES
P02751	P24821	0.99999923	Known	YES	NO	YES	YES
P07996	Q07954	0.99999923	Known	YES	YES	NO	YES
P11277	Q13813	0.99999923	Known	YES	YES	YES	YES
P54259	P98160	0.99999922	Known	NO	NO	NO	YES
O15273	Q8WZ42	0.99999921	Known	YES	YES	YES	YES
P98095	P98160	0.99999921	Known	YES	NO	NO	YES
P17301	P98160	0.99999921	Known	NO	NO	YES	YES
P02549	P11277	0.99999921	Known	YES	YES	YES	YES
Q99698	Q9HC77	0.9999992	Known	YES	NO	NO	YES
Q01484	Q92823	0.9999992	Known	YES	YES	NO	NO
Q13535	Q14839	0.9999992	Known	YES	YES	NO	YES
Q14676	Q92793	0.9999992	Known	YES	NO	NO	YES
P01024	P25391	0.9999992	Known	YES	YES	NO	NO
Q14686	Q92793	0.99999919	Known	YES	NO	NO	YES
Q09472	Q14676	0.99999919	Known	YES	YES	YES	YES
P51587	Q9BXW9	0.99999919	Known	YES	YES	YES	YES
P16157	Q14573	0.99999919	Known	YES	NO	NO	NO
Q96K83	Q99698	0.99999918	Known	NO	NO	NO	NO
Q12873	Q5UIP0	0.99999918	Known	YES	NO	NO	YES
Q09472	Q14686	0.99999918	Known	YES	NO	YES	YES
P11532	Q9ULU8	0.99999918	Known	YES	YES	NO	NO
O15083	Q9Y6V0	0.99999917	Known	YES	NO	NO	NO
P38398	Q13535	0.99999917	Known	YES	NO	YES	YES
P32004	Q01484	0.99999917	Known	YES	YES	NO	NO
P02462	Q02388	0.99999917	Known	YES	NO	NO	YES
O95477	Q9Y566	0.99999917	Known	YES	NO	NO	NO
Q96RY7	Q9NYQ7	0.99999916	Known	NO	NO	NO	NO
P18206	P98161	0.99999916	Known	NO	NO	NO	YES
O60674	P98161	0.99999916	Known	YES	NO	NO	YES
Q8IWT3	Q9ULL4	0.99999915	Known	NO	NO	NO	NO
Q71F56	Q9UHV7	0.99999915	Known	YES	YES	YES	YES
Q6P0Q8	Q92616	0.99999915	Known	YES	YES	NO	NO

... (continued)

Appendix 4-4 : Number interacting pairs where both proteins have the same function (A) or are involved in the same cellular process (B) for previously reported data compared with those reported in this study only.

Appendix 4-5: List of interactions for fibroblast growth factors (FGFs) and cyclin-dependent kinases (CDKs) and FGF receptors and CDK inhibitor/activators. * are interactors that have not been previously reported.

The interactions of fibroblast growth factors (FGFs) with fibroblast growth factor receptors (FGFRs).	
FGF	FGFR
P61328	
P31371	
Q9NSA1	
Q9GZV9	
P21781	P11362*, P22607*, P21802
O43320	
O95750	P22455
O15520	P21802
P11487	
P55075	
P12034	
P10767	
P08620	
P09038	P11362, P22607*, P21802*
Q92914	
Q92915	
O76093	
Q92913	
P05230	P11362, P22607, P21802, P22455
O60258	
Q9HCT0	
Q9NP95	
The interactions of cyclin-dependent kinases (CDKs) with regulators of cyclin-dependent kinases.	
CDK	Regulator of CDK
Q8IZL9	
Q96Q40	P63104*, P31946*
O94921	P63104*, P31946
P50613	
Q9UQ88	P31946*, P63104*
Q00526	P31946*, P49918*, P42773*, Q16667, P46527, P38936, P55273*, Q13319*, P63104*, Q15078*, Q96SZ6*, Q8IWU2*, O15392*
Q00534	O15392*, P42772*, P46527, P49918*, P42773, P38936, Q8N726, P55273
Q00537	P63104, P31946*, Q15078*
Q00536	P63104, P31946*, Q15078
Q00535	O15392*, P38936*, P46527, P31946*, Q13319, Q96SZ6, Q8IWU2, P63104*, Q15078, Q16667*
Q9BWU1	O75448*, P42772*, Q15648*, Q93074*, Q9UHV7*, O75586*
Q14004	
Q9NYV4	
Q15131	P31946*, P63104*
P50750	Q15078
P06493	P46527*, Q13319*, P63104*, Q96SZ6*, P38936, Q8IWU2*, P31946*, Q16667, O15392, Q15078*
P11802	P38936, P42773, Q8N726, P55273, P42772, P46527, P49918, O15392, P42771*
P21127	P31946, P63104*
P24941	P38936, P31946*, P63104*, P46527, Q16667, Q8IWU2*, P49918*, Q15078*, Q96SZ6*, Q13319*, O15392
Q07002	P31946*, P63104*, Q15078*
P49336	O75586, P42771*, Q93074, Q15648, P42772, Q9UHV7, O75448

Appendix 4-6: List of proteins from the acute phase response pathway, complement pathway and glucocorticoids receptor pathway.

A1L4G7	O75051	P04150	P19419	P63165	Q5JNX2	Q9UGI4
A5PL27	O75376	P04196	P19875	P63279	Q5SBK4	Q9UMI2
B1AMY1	O95608	P05112	P21730	P81172	Q5T7S2	Q9Y478
B3KNT3	P00736	P05113	P22301	Q00403	Q6I9T4	Q9Y6K9
B5BUQ7	P00738	P05121	P23458	Q02790	Q6LDG4	Q9Y6Q9
C9IYG8	P00746	P05231	P25963	Q04206	Q6LDI0	
C9IZP8	P00749	P05412	P27352	Q04917	Q6LDP0	
C9J1C7	P00751	P05546	P28562	Q06AH7	Q6LE88	
C9J1D9	P01009	P06401	P29353	Q06DL9	Q6NWP6	
C9JC72	P01011	P06681	P31749	Q07817	Q6PIW7	
C9JHD2	P01019	P07550	P32241	Q09472	Q6QR78	
C9JU00	P01023	P07900	P34932	Q13233	Q86SI0	
C9JV77	P01024	P08603	P35225	Q13451	Q8TCE8	
D3DUU2	P01031	P08700	P35228	Q13514	Q8TCF0	
D5M8Q2	P01100	P08887	P38936	Q13546	Q8TD58	
D6REL8	P01160	P09429	P40424	Q14551	Q92831	
E7EMZ6	P01375	P09601	P42224	Q14624	Q96DZ4	
E7ERA4	P01579	P10145	P42345	Q14978	Q99557	
E7ET33	P01584	P10147	P45983	Q14UF5	Q99558	
E9PD65	P02735	P10415	P48552	Q15185	Q99576	
E9PDS4	P02741	P10643	P48736	Q15596	Q99616	
E9PER2	P02743	P11226	P49715	Q15628	Q99650	
E9PF72	P02749	P11684	P49918	Q15648	Q99683	
E9PI80	P02760	P12314	P51606	Q15750	Q99836	
E9PJN2	P02763	P13500	P51617	Q15788	Q99893	
E9PK97	P02766	P13501	P51692	Q16334	Q99933	
O00187	P02790	P15529	P52333	Q16581	Q9BQ95	
O14543	P02818	P16581	P61201	Q16822	Q9BXH2	
O43524	P04049	P17676	P61978	Q19MP7	Q9BYG6	
O60244	P04083	P18510	P62993	Q4FCH6	Q9NZ70	
O60424	P04141	P19320	P63000	Q506Q0	Q9UGE8	

Appendix 4-7: List of hub proteins in *Homo sapiens* ranked by the number of neighbors for each protein.

Protein	# of Neighbours	# of Known Neighbours	# of Novel Neighbours
Q156A1	3773	0	3773
P02811	1174	0	1174
Q15072	1018	0	1018
P51522	1004	2	1002
Q5EBM4	938	0	938
Q96CX3	937	0	937
Q8N782	895	0	895
Q3ZCT1	884	0	884
Q8N3J9	875	0	875
P84022	836	184	652
Q96N58	797	0	797
Q96SK3	759	1	758
Q9H963	743	0	743
Q14585	741	0	741
P31946	722	127	595
P63104	721	135	586
P17039	721	0	721
Q04917	719	60	659
P27348	716	114	602
P61981	713	245	468
P31947	710	108	602
Q8NB50	706	0	706
Q8TAQ5	694	0	694
Q15796	672	168	504
P62258	671	70	601
Q8TF20	671	0	671
P12931	670	223	447
Q99717	669	37	632
Q15797	661	115	546
Q6ZR52	645	0	645
Q15940	634	0	634
P06241	624	153	471
P07948	624	100	524
P06239	615	103	512
O15198	611	105	506
Q6AZW8	604	0	604
Q6ZTB9	600	0	600
P08631	597	53	544
P09769	594	18	576
A8MUV8	590	0	590

P07947	584	29	555
Q6P3V2	583	0	583
Q52M93	582	1	581
Q9Y2P7	581	0	581
Q9H7R5	578	0	578
A6NN14	568	0	568
P51451	554	12	542
Q6ZN57	545	1	544
P04637	527	263	264
Q8N972	520	0	520
Q3KP31	516	0	516
A8MXY4	504	0	504
O43345	495	1	494
Q05481	494	1	493
P17252	493	173	320
P06493	481	120	361
P05771	477	69	408
P28482	470	168	302
Q9UII5	468	2	466
Q5SXM1	467	0	467
P27361	465	116	349
Q09472	464	219	245
Q9P2J8	460	0	460
Q92793	454	219	235
P68133	452	111	341
P63267	451	2	449
P63261	451	46	405
P60709	451	106	345
Q562R1	450	0	450
P68032	450	32	418
Q9BYX7	449	0	449
P62736	449	15	434
P10275	446	158	288
Q13485	445	154	291
P03372	437	203	234
P24941	436	95	341
Q86YE8	435	0	435
Q00526	428	15	413
P05129	425	47	378
P38398	424	161	263
P17017	419	0	419

Q6ZNH5	395	0	395
Q15834	392	125	267
P06400	386	145	241
Q13547	382	151	231
P62993	372	217	155
P17030	371	0	371
Q00535	367	49	318
Q92769	364	91	273
Q8TBZ8	364	0	364
P17612	357	158	199
P22694	355	4	351
Q99750	352	101	251
O15379	351	68	283
P04280	351	0	351
P22612	350	1	349
Q8N7K0	340	0	340
Q9BYQ5	339	0	339
Q9BYQ8	338	0	338
Q9BYQ6	338	0	338
Q9BYR0	338	0	338
Q9BYR3	338	0	338
Q9BQ66	338	74	264
A8MXZ3	338	0	338
Q9BYR4	338	0	338
Q9BYR5	338	0	338
Q14591	333	0	333
Q9BYQ7	332	0	332
Q15929	331	0	331
Q9BYQ9	331	0	331
Q9BYR2	331	0	331
A8MTL4	331	0	331
P63000	329	101	228
Q6ZN19	326	0	326
P61163	325	6	319
Q96IR2	324	0	324
Q13588	322	11	311
Q13263	319	26	293
P84095	317	18	299
Q9NRR5	316	143	173
P68400	313	175	138
P19784	311	76	235

P00533	310	159	151
P04281	308	0	308
Q8TC17	307	0	307
Q9BYD9	305	0	305
P60763	304	5	299
Q6VMQ6	303	61	242
P15622	303	11	292
O14964	300	89	211
P36897	299	147	152
Q08AN1	297	0	297
Q9BX82	297	0	297
O43830	296	0	296
Q12933	296	148	148
Q01844	295	115	180
Q9NXT0	294	0	294
P63167	294	64	230
Q96FJ2	294	11	283
P15153	292	17	275
Q9NSJ1	292	0	292
Q8N8L2	291	0	291
Q9Y4K3	291	105	186
P24928	288	84	204
Q96JF6	286	1	285
Q15973	283	0	283
P08670	280	111	169
B1ANC4	280	0	280
Q9Y6R6	279	0	279
P54253	278	146	132
P17038	271	0	271
P31749	271	121	150
P63165	270	66	204
P35222	268	146	122
Q8IYB9	267	0	267
O75290	266	0	266
P31751	266	20	246
P61956	265	29	236
P61586	263	84	179
Q7L3S4	263	0	263
Q96I27	263	0	263
P60953	261	96	165
Q9Y2A4	261	0	261

Q05516	260	83	177
Q6EEV6	259	74	185
Q96GE5	258	0	258
Q5T3J3	257	95	162
P56524	255	49	206
Q6ZNG1	255	0	255
Q5UIP0	254	89	165
P08134	253	6	247
P17081	251	17	234
O14641	249	50	199
P04150	249	99	150
Q86YH7	248	0	248
P52743	247	0	247
P49910	244	14	230
Q494X3	244	0	244
Q13432	243	79	164
O95129	242	0	242
P51504	242	0	242
P42025	241	3	238
P11142	241	66	175
P55854	240	20	220
O95132	240	0	240
Q4V348	240	0	240
Q86UE3	239	0	239
Q9Y243	239	5	234
Q6ZNA1	239	0	239
P14136	238	23	215
P08107	238	7	231
P54652	238	4	234
O15162	237	71	166
Q8N8Z8	237	0	237
Q15505	236	0	236
P35789	236	0	236
P34931	235	0	235
P16415	235	0	235
P54257	233	56	177
P17024	233	2	231
Q6N045	232	0	232

... (continued)

Appendix 4-8: List of betweenness centrality proteins in *Homo sapiens* ranked by highest betweenness centrality measure.

Protein	Betweenness Centrality								
Q156A1	52021233.9	P05129	626372.9027	P78352	383536.6502	Q9UNE7	287977.0422	P61163	236059.9985
P02811	4789053.495	P07947	622006.694	Q15047	374974.569	Q14055	286652.2885	P21673	236009.3696
P84022	2020362.837	P68400	615077.4797	P29350	369556.5808	P40337	285680.4777	P67870	234386.0056
P04637	1512443.906	P19784	610290.9836	Q14585	359840.7075	Q92793	285540.8038	Q57313	233453.1176
Q01844	1502425.216	P51451	602386.2418	P46379	354102.8926	Q13873	285321.1953	P17661	230641.411
Q9NRR5	1308161.975	P03372	599551.7374	P63279	351720.3849	Q5UIP0	284660.9884	P11142	229619.9209
Q04917	1081315.053	P17612	592693.4733	P14136	347096.1649	Q07954	284568.7354	P36873	228331.2813
P31946	1080073.719	P22694	588460.7496	Q96N58	346680.7353	Q99972	281665.9663	Q12959	228114.5168
P63104	1074074.893	P35222	577673.2005	P05067	346670.8886	Q09472	280564.4606	P01042	228091.1253
P61981	1066679.146	P12956	568535.4678	Q06124	346652.7926	P17039	279781.5959	P02647	227794.9739
P27348	1055129.145	P22612	568162.7053	P31749	342284.2401	P63208	278163.6834	P09958	227183.2269
P31947	1020967.678	Q00535	562788.8777	Q8NER5	341501.3381	P40763	276376.1561	Q12873	226141.6866
Q15796	997823.9529	Q3ZCT1	539780.0972	Q13547	339785.8519	O15379	275066.6071	Q8N726	226070.1162
P12931	968522.5928	P42574	532473.4279	P56537	325428.6134	Q07817	273685.1291	P08107	223197.6805
P62258	936053.7736	Q8N782	526007.7482	O95777	323741.9206	P10275	272804.2886	P54652	223197.6805
P06493	924389.88	P21246	499711.1552	P63000	323535.8183	P55854	272671.3297	Q6ZN57	222390.0087
Q15072	905510.9073	P28482	489364.9159	Q15645	322672.9701	O75400	270424.6946	P36575	221710.3069
P17252	896109.4058	Q6VMQ6	475817.4288	P63165	322653.9224	P56539	267258.7326	P21333	221273.0424
Q99717	889536.9031	Q8N3J9	473968.6411	P31751	321530.7232	P63167	266021.0142	Q99689	220979.2807
P51522	856438.3568	Q9UMX0	470085.947	P10415	320624.7273	Q96FJ2	266021.0142	P62136	220717.7876
Q15797	844447.963	P04278	469193.0433	O15360	317689.1556	O9Y243	265311.7281	Q6AZW8	218180.2579
Q15834	837591.8195	P27361	462853.3575	P02751	317480.6509	O14745	264454.6557	P24928	217615.0626
P07948	836470.2039	Q13432	441731.4494	P54257	315131.2275	Q06609	263866.3293	Q9NZR2	217034.5122
P00533	832163.2601	P06400	439611.8266	P61956	315078.9184	Q8TC17	262492.4542	O14641	216804.2392
P05771	810413.1646	P68133	432407.786	Q7LSN1	310155.6736	Q92796	262133.5964	Q16539	216133.3308
O15198	778728.1743	Q8TAQ5	417345.8298	P04280	310094.6182	Q15287	261679.1612	P34931	215554.7688
Q12933	777884.9436	P14923	415283.5588	Q13588	309839.1937	P01106	260629.7906	P20023	210689.6985
P24941	776428.9543	P68032	415173.6732	P68104	306817.9115	P60763	259338.3746	Q9BYD9	207897.8158
P36897	773216.1677	P60709	414821.7951	Q05639	306817.9115	Q9H963	257228.1751	Q9Y6T7	207161.6667
Q00526	758712.9459	P63261	414821.7951	Q5VTE0	306817.9115	P04049	256666.8299	P54253	207067.6117
P38398	744745.8584	P63267	414821.7951	Q14050	305987.7037	Q9H7R5	256504.3926	O15162	205845.8895
P06241	739865.7993	P23743	414390.8148	P08195	305488.2983	Q03135	254665.5215	Q9Y3C5	205781.1239
Q13485	735311.963	Q562R1	411180.115	Q96SK3	305018.1426	Q16665	252528.0956	Q9NRY7	203812.8736
P06239	729641.8094	P62736	409673.9766	P04281	304859.9121	Q9NY93	252510.8494	O15287	203196.2827
Q96CX3	680181.1306	Q9BYX7	408689.39	P08123	302179.7905	P22607	249439.2376	Q15700	203070.0635
Q99750	679463.9587	P05556	407941.0009	Q6EEV6	300214.0731	Q8NB50	247371.4905	P62140	202601.377
P08631	676504.3603	P12757	406611.3195	Q05655	299604.8998	P15153	245462.5229	P01111	201076.925
P09769	654735.3606	Q9HAU4	405209.0784	Q9HCE7	299214.8227	Q8TF20	245224.6769		
Q9Y4K3	631413.5406	P62993	403615.1116	Q92769	296664.9502	P01112	244979.9036		
Q5EBM4	629314.483	P08670	401341.9222	P84095	295983.9751	O43586	244829.4337		
		P26641	398500.007	P00519	293204.22	Q14721	238713.0272		

... (continued)

Appendix 4-9: List of protein mutations associated with resistance to breast cancer therapeutics doxorubicin and Trastuzumab.

Mutation Name	Protein Name	Mutation
P04637a	P04637	First 204 amino acids only
P04637b	P04637	Codon 249, R replaced with G
P04637c	P04637	Codons 217-221 deleted
P04637d	P04637	Codon 248, R replaced with Q
P04626a	P04626	First 611 amino acids only
P04626b	P04626	First 648 amino acids only

Part B

List of publications

1: **Samanfar B**, Tan le Hoa, Shostak K, Chalabian F, Wu Z, Alamgir M, Sunba N, Burnside D, Omid K, Hooshyar M, Galván Márquez I, Jessulat M, Smith ML, Babu M, Azizi A, Golshani A. A global investigation of gene deletion strains that affect premature stop codon bypass in yeast, *Saccharomyces cerevisiae*. Mol Biosyst. 2014, 10(4):916-924.

Molecular
BioSystems

PAPER

[View Article Online](#)
[View Journal](#)

A global investigation of gene deletion strains that affect premature stop codon bypass in yeast, *Saccharomyces cerevisiae*†

Cite this: DOI: 10.1039/c3mb70501c

Bahram Samanfar,^a Le Hoa Tan,^a Kristina Shostak,^a Firoozeh Chalabian,^{ab} Zongbin Wu,^a Md Alamgir,^a Noor Sunba,^a Daniel Burnside,^a Katayoun Omid K,^a Mohsen Hooshyar,^a Imelda Galván Márquez,^a Matthew Jessulat,^{ac} Myron L. Smith,^a Mohan Babu,^c Ali Azizi^d and Ashkan Golshani^{*a}

Protein biosynthesis is an orderly process that requires a balance between rate and accuracy. To produce a functional product, the fidelity of this process has to be maintained from start to finish. In order to systematically identify genes that affect stop codon bypass, three expression plasmids, pUKC817, pUKC818 and pUKC819, were integrated into the yeast non-essential loss-of-function gene array (5000 strains). These plasmids contain three different premature stop codons (UAA, UGA and UAG, respectively) within the *LacZ* expression cassette. A fourth plasmid, pUKC815 that carries the native *LacZ* gene was used as a control. Transformed strains were subjected to large-scale β -galactosidase lift assay analysis to evaluate production of β -galactosidase for each gene deletion strain. In this way 84 potential candidate genes that affect stop codon bypass were identified. Three candidate genes, *OLA1*, *BSC2*, and *YNL040W*, were further investigated, and were found to be important for cytoplasmic protein biosynthesis.

Received 11th November 2013,
Accepted 8th February 2014

DOI: 10.1039/c3mb70501c

www.rsc.org/molecularbiosystems

2: **Samanfar B**, Omid K, Hooshyar M, Laliberte B, Alamgir M, Seal AJ, Ahmed-Muhsin E, Viteri DF, Said K, Chalabian F, Golshani A, Wainer G, Burnside D, Shostak K, Bugno M, Willmore WG, Smith ML, Golshani A. Large-scale investigation of oxygen response mutants in *Saccharomyces cerevisiae*. *Mol Biosyst*. 2013, 9(6):1351-1359.

Large-scale investigation of oxygen response mutants in *Saccharomyces cerevisiae*[†]

Cite this: DOI: 10.1039/c3mb25516f

Bahram Samanfar,^{ab} Katayoun Omid,^{ab} Mohsen Hooshyar,^{ab} Ben Laliberte,^{ab} MD Alamgir,^{ab} Andrew J. Seal,^{ab} Eman Ahmed-Muhsin,^{ab} Duber Frey Viteri,^{ab} Kamaleldin Said,^{ac} Firoozeh Chalabian,^{ad} Ardeshir Golshani,^e Gabriel Wainer,^e Daniel Burnside,^{ab} Kristina Shostak,^{ab} Magdalena Bugno,^{ab} William G. Willmore,^{ab} Myron L. Smith^{ab} and Ashkan Golshani^{*ab}

A genome-wide screen of a yeast non-essential gene-deletion library was used to identify sick phenotypes due to oxygen deprivation. The screen provided a manageable list of 384 potentially novel as well as known oxygen responding (anoxia-survival) genes. The gene-deletion mutants were further assayed for sensitivity to ferrozine and cobalt to obtain a subset of 34 oxygen-responsive candidate genes including the known hypoxic gene activator, *MGA2*. With each mutant in this subset a plasmid based β -galactosidase assay was performed using the anoxic-inducible promoter from *OLE1* gene, and 17 gene deletions were identified that inhibit induction under anaerobic conditions. Genetic interaction analysis for one of these mutants, the RNase-encoding *POP2* gene, revealed synthetic sick interactions with a number of genes involved in oxygen sensing and response. Knockdown experiments for *CNOT8*, human homolog of *POP2*, reduced cell survival under low oxygen condition suggesting a similar function in human cells.

Received 12th November 2012,
Accepted 22nd February 2013

DOI: 10.1039/c3mb25516f

www.rsc.org/molecularbiosystems

3: Pitre S, Hooshyar M, Schoenrock A, **Samanfar B**, Jessulat M, Green JR, Dehne F, Golshani A. Short Co-occurring Polypeptide Regions Can Predict Global Protein Interaction Maps. Sci Rep. 2012, 2:239.

SUBJECT AREAS:
BIOINFORMATICS
SYSTEMS BIOLOGY
COMPUTATIONAL BIOLOGY
GENOMICS

Received
2 November 2011

Accepted
14 December 2011

Published
30 January 2012

Correspondence and
requests for materials
should be addressed to

Short Co-occurring Polypeptide Regions Can Predict Global Protein Interaction Maps

Sylvain Pitre¹, Mohsen Hooshyar², Andrew Schoenrock¹, Bahram Samanfar², Matthew Jessulat², James R. Green³, Frank Dehne¹ & Ashkan Golshani²

¹School of Computer Science, Carleton University, Ottawa, Canada. ²Department of Biology, Carleton University, Ottawa, Canada. ³Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada.

A goal of the post-genomics era has been to elucidate a detailed global map of protein-protein interactions (PPIs) within a cell. Here, we show that the presence of co-occurring short polypeptide sequences between interacting protein partners appears to be conserved across different organisms. We present an algorithm to automatically generate PPI prediction method parameters for various organisms and illustrate that global PPIs can be predicted from previously reported PPIs within the same or a different organism using protein primary sequences. The PPI prediction code is further accelerated through the use of parallel multi-core programming, which improves its usability for large scale or proteome-wide PPI prediction. We predict and analyze hundreds of novel human PPIs, experimentally confirm protein functions and importantly predict the first genome-wide PPI maps for *S. pombe* (~9,000 PPIs) and *C. elegans* (~37,500 PPIs).

4: Babu M, Arnold R, Bundalovic-Torma C, Gagarinova A, Wong KS, Kumar A, Stewart G, **Samanfar B**, Aoki H, Wagih O, Vlasblom J, Phanse S, Lad K, Yeou Hsiung Yu A, Graham C, Jin K, Brown E, Golshani A, Kim P, Moreno-Hagelsieb G, Greenblatt J, Houry WA, Parkinson J, Emili A. Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genet*. 2014, 10(2):e1004120.

OPEN ACCESS Freely available online

PLOS | GENETICS

Quantitative Genome-Wide Genetic Interaction Screens Reveal Global Epistatic Relationships of Protein Complexes in *Escherichia coli*

Mohan Babu^{1,2,3*}, Roland Arnold^{1,3}, Cedoljub Bundalovic-Torma^{3,4,5}, Alla Gagarinova^{1,5,6}, Keith S. Wong^{4,6}, Ashwani Kumar², Geordie Stewart⁶, Bahram Samanfar⁷, Hiroyuki Aoki², Omar Wagih¹, James Vlasblom², Sadhna Phanse^{1,2}, Krunal Lad², Angela Yeou Hsiung Yu⁴, Christopher Graham², Ke Jin^{1,2}, Eric Brown⁶, Ashkan Golshani⁷, Philip Kim¹, Gabriel Moreno-Hagelsieb⁸, Jack Greenblatt^{1,5}, Walid A. Houry⁴, John Parkinson^{3,4,5}, Andrew Emili^{1,5*}

¹ Banting and Best Department of Medical Research, Donnelly Centre, University of Toronto, Toronto, Ontario, Canada, ² Department of Biochemistry, Research and Innovation Centre, University of Regina, Regina, Saskatchewan, Canada, ³ Hospital for Sick Children, Toronto, Ontario, Canada, ⁴ Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada, ⁵ Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada, ⁶ Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada, ⁷ Department of Biology and Ottawa Institute of Systems Biology, Carleton University, Ottawa, Ontario, Canada, ⁸ Department of Biology, Wilfrid Laurier University, Waterloo, Ontario, Canada

Abstract

Large-scale proteomic analyses in *Escherichia coli* have documented the composition and physical relationships of multiprotein complexes, but not their functional organization into biological pathways and processes. Conversely, genetic interaction (GI) screens can provide insights into the biological role(s) of individual gene and higher order associations. Combining the information from both approaches should elucidate how complexes and pathways intersect functionally at a systems level. However, such integrative analysis has been hindered due to the lack of relevant GI data. Here we present a systematic, unbiased, and quantitative synthetic genetic array screen in *E. coli* describing the genetic dependencies and functional cross-talk among over 600,000 digenic mutant combinations. Combining this epistasis information with putative functional modules derived from previous proteomic data and genomic context-based methods revealed unexpected associations, including new components required for the biogenesis of iron-sulphur and ribosome integrity, and the interplay between molecular chaperones and proteases. We find that functionally-linked genes co-conserved among γ -proteobacteria are far more likely to have correlated GI profiles than genes with divergent patterns of evolution. Overall, examining bacterial GIs in the context of protein complexes provides avenues for a deeper mechanistic understanding of core microbial systems.

5: Omid K, Hooshyar M, Jessulat M, **Samanfar B**, Sanders M, Burnside D, Pitre S, Schoenrock A, Xu J, Babu M, Golshani A. Phosphatase complex Pph3/Psy2 is involved in regulation of efficient non-homologous end-joining pathway in the yeast *Saccharomyces cerevisiae*. PLoS One. 2014, 9(1):e87248.

OPEN ACCESS Freely available online

Phosphatase Complex Pph3/Psy2 Is Involved in Regulation of Efficient Non-Homologous End-Joining Pathway in the Yeast *Saccharomyces cerevisiae*

Katayoun Omid^{1,2}, Mohsen Hooshyar^{1,2}, Matthew Jessulat^{1,3}, Bahram Samanfar^{1,2}, Megan Sanders^{1,2}, Daniel Burnside^{1,2}, Sylvain Pitre⁴, Andrew Schoenrock⁴, Jianhua Xu⁵, Mohan Babu³, Ashkan Golshani^{1,2*}

¹ Department of Biology, Carleton University, Ottawa, Ontario, Canada, ² Ottawa Institute of Systems Biology, Carleton University, Ottawa, Ontario, Canada, ³ Department of Biochemistry, Research and Innovation Centre, University of Regina, Regina, Saskatchewan, Canada, ⁴ Department of Computer Science, Carleton University, Ottawa, Ontario, Canada, ⁵ College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang, China

Abstract

One of the main mechanisms for double stranded DNA break (DSB) repair is through the non-homologous end-joining (NHEJ) pathway. Using plasmid and chromosomal repair assays, we showed that deletion mutant strains for interacting proteins Pph3p and Psy2p had reduced efficiencies in NHEJ. We further observed that this activity of Pph3p and Psy2p appeared linked to cell cycle Rad53p and Chk1p checkpoint proteins. Pph3/Psy2 is a phosphatase complex, which regulates recovery from the Rad53p DNA damage checkpoint. Overexpression of Chk1p checkpoint protein in a parallel pathway to Rad53p compensated for the deletion of *PPH3* or *PSY2* in a chromosomal repair assay. Double mutant strains $\Delta pph3/\Delta chk1$ and $\Delta psy2/\Delta chk1$ showed additional reductions in the efficiency of plasmid repair, compared to both single deletions which is in agreement with the activity of Pph3p and Psy2p in a parallel pathway to Chk1p. Genetic interaction analyses also supported a role for Pph3p and Psy2p in DNA damage repair, the NHEJ pathway, as well as cell cycle progression. Collectively, we report that the activity of Pph3p and Psy2p further connects NHEJ repair to cell cycle progression.

6: Jessulat M, Pitre S, Gui Y, Hooshyar M, Omidi K, **Samanfar B**, Tan le Hoa, Alamgir M, Green J, Dehne F, Golshani A. Recent advances in protein-protein interaction prediction: experimental and computational methods. *Expert Opin Drug Discov.* 2011, 6(9):921-935.

Review

Expert Opinion

1. Introduction
2. Two-hybrid methods
3. Pull-down methods
4. Protein microarrays
5. Computational and predictive methods of PPI analysis
6. The protein interaction prediction engine (PIPE)
7. Evaluation of PPI prediction methods
8. Discussions
9. Expert opinion

Recent advances in protein-protein interaction prediction: experimental and computational methods

Matthew Jessulat, Sylvain Pitre, Yuan Gui, Mohsen Hooshyar, Katayoun Omidi, Bahram Samanfar, Le Hoa Tan, Md Alamgir, James Green, Frank Dehne & Ashkan Golshani[†]

[†]*Carleton University, Department of Biology, Ottawa, Ontario, Canada*

Introduction: Proteins within the cell act as part of complex networks, which allow pathways and processes to function. Therefore, understanding how proteins interact is a significant area of current research.

Areas covered: This review aims to present an overview of key experimental techniques (yeast two-hybrid, tandem affinity purification and protein microarrays) used to discover protein-protein interactions (PPIs), as well as to briefly discuss certain computational methods for predicting protein interactions based on gene localization, phylogenetic information, 3D structural modeling or primary protein sequence data. Due to the large-scale applicability of primary sequence-based methods, the authors have chosen to focus on this strategy for our review. There is an emphasis on a recent algorithm called Protein Interaction Prediction Engine (PIPE) that can predict global PPIs. The readers will discover recent advances both in the practical determination of protein interaction and the strategies that are available to attempt to anticipate interactions without the time and costs of experimental work.

Expert opinion: Global PPI maps can help understand the biology of complex diseases and facilitate the identification of novel drug target sites. This study describes different techniques used for PPI prediction that we believe will significantly impact the development of the field in a new future. We expect to see a growing number of similar techniques capable of large-scale PPI predictions.