

**Prediction and investigation of novel proteins in DNA double
stranded break repair**

by

Mohsen Hooshyar

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

Biology

Department of Biology
Ottawa-Carleton Institute of Biology
Carleton University
Ottawa, Ontario

© 2015, Mohsen Hooshyar

Abstract

DNA double stranded breaks (DSBs) are the most genotoxic forms of DNA lesions, causing fragmentation of the DNA strands. Mis-repaired and unrepaired DSBs lead to chromosomal rearrangement and genomic instability promoting tumorigenesis or cell death. DSBs are primarily repaired by two independent and highly conserved pathways: homologous recombination (HR) and non-homologous end joining (NHEJ). HR requires a homologous sequence to repair DNA breaks, whereas NHEJ repair is achieved through direct ligation of the broken ends of DNA. The process of NHEJ involves three main protein complexes: Yku70/Yku80 initiates and stabilizes the DNA ends, Mre11/Rad50/Xrs2 brings broken ends to close proximity and Dnl4/Lif1/Nej1 ligates the DNA ends. Protein-protein interaction (PPI) has been utilized in functional genomics studies to identify novel proteins involved in different pathways based on their PPI profiles. In this study we aim to screen for novel proteins involved in NHEJ using PPI predictions. We use a computational tool to predict novel PPIs between DNA repair proteins and human proteome. Using this method, we predicted 271 novel PPIs, expanding the reported human DNA repair interactome by 75%. Yeast homologs of the novel human gene candidates were subjected to a plasmid-based repair assay, in which deletion of eight of 12 yeast genes showed reduction in the repair efficiency. We further investigated the roles of Tpk1, Arp6 and Psk1/Psk2 in NHEJ repair. Deletion of any one of *TPK1*, *ARP6*, *PSK1* and *PSK2* reduced efficiency of repair in chromosomal and plasmid based assays and showed sensitivity to different DNA damaging agents. We suggest that Tpk1, catalytic subunit of protein kinase A, is involved in recruitment of Nej1 to the site of damage and its role in NHEJ is dependent on Nej1. We propose that the involvement of actin related protein 6, Arp6, in NHEJ is linked to Rsc2 and

Mre11. Additionally, we suggest a role for PAS kinase complex, Psk1/Psk2, that is connected to DNA damage checkpoint proteins.

Acknowledgments

First and foremost, I would like to thank Dr. Golshani for giving me the opportunity to work in his lab. I would like to express my gratitude for his support, guidance and mentorship. But mostly thanks for your patience and believing in me, and for helping me to reach goals that would've not been possible without your supervision, inspiration and encouragement.

I would like to thank Dr. Rudner and Dr. Smith for their critical inputs and contribution. It was a great opportunity to have Dr. Smith on my committee and to work for him as a TA, as he provided invaluable advice and feedback regarding my research. I wish to extend my appreciation to Dr. Babu and his lab, specially Matt, for their willingness to collaborate. I would also like to thank our bioinformatics team and collaborators Dr. Dehne, Dr. Green, Dr. Pitre and Andrew for their help in developing a powerful prediction tool. I would like to thank Suzzanne Patterson for being so helpful and accomedating. I would like to thank Biology staff, Laura, Caitlyn, Ruth, Ed and Darlene for being so helpful.

To all the veteran members of Golshani lab, Noor, Matt, Alam and Nazila, thanks for welcoming me into the lab and patiently teaching me the alphabets of lab work. To the G-lab greats Bahram, Katie, Dan, Maryam, Kristina, Yuan, Imelda, Houman and Le Hoa, it was a joy to work with you. Thanks for all your help, support and friendship. I have been fortunate to have the best undergraduate students through the course of my PhD, Alex, Megan, Adamo, Pedro, Urvi and Ramy, I appreciate all the hard work. To the graduate students, Magda, Ally, Duale, Eddie, Matt, Mary, Shana, AJ and Haiyun thanks for creating a friendly environment.

I want to specially thank, Magda for her love, care and support, Ally for always being there, Bahram and Katie for their kindness and big heart, Dan and Maryam for their help and friendship. You made this a memorable journey. I also want to thank my closest friends, Adel, Amir, Farhad, Farzad, Foroutan, Mamad and Sayeh.

Last but not least, I would like to express my sincere gratitude and appreciation to my family, specially my parents. Thank you for your unconditional love and support, and for providing everything possible to help me achieve my goals, as I could not be where I am without you. I would like to dedicate this work to my wonderful parents.

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

Statement of contribution

Chapter 2, “Short Co-occurring Polypeptide Regions Can Predict Global Protein Interaction Maps”. I participate in the design and development of the project and manuscript preparation. I was in charge of data collection for human PPI. Computational work was done by Dr. Dehne and Dr. Green’s group in computer sciences and engineering department. Analysis and evaluation of collected data, identification of novel complexes in *S. pombe* and novel protein targets in human was done by me. Wet lab experiments for this work were performed by me. B. Samanfar and M. Jessulat helped with data analysis.

Chapter 3, “TPK1 influences the efficiency of non-homologous end joining “. Hypothesis, experimental design, development and data analysis were done by me. Experiments in yeast cell were carried out by me. Dr. Babu’s lab from University of Regina carried out the mammalian work in this chapter. A. Jesso helped with plasmid repair assay.

Chapter 4, “Actin Related Protein 6, Arp6, is involved in repair of DNA double stranded break in *Saccharomyces cerevisiae* “. I came up with the hypothesis, experimental design and development of the project. D. Burnside and A. Jesso helped with plasmid repair assay and SGA analysis.

Chapter 5, “PAS kinase complex regulates the efficiency and fidelity of non-homologous end joining repair”. I designed the experiments and developed the project. I carried out the experiments for Psk2 and D. Burnside carried out the experiments for Psk1. I and D. Burnside analyzed the data obtained from SGA and PSA.

Table of contents

Abstract	ii
Acknowledgments	iv
Statement of contribution	vi
Table of content	vii
List of tables	xiii
List of figures	xiv
List of appendices	xvi
List of abbreviations	xviii
Chapter 1: Introduction	1
1.1 DNA damage	1
1.1.1 DNA double stranded breaks repair	3
1.1.2 DNA damage response	5
1.1 Homologous recombination	6
1.1 Non-homologous end joining	11
1.2 System biology	15
1.2.1 Functional genomics	17
1.2.2 Genetic interaction	19
1.2.3 Protein-protein interaction	22
1.3 Focus and objectives	26

Chapter 2: Short co-occurring polypeptide regions can predict global protein-protein interaction maps	29
2.1 Abstract	29
2.2 Introduction	29
2.3 Methods	31
2.3.1 PPI prediction via interaction codes	31
2.3.2 PPI prediction adaption algorithm	32
2.3.3 Improved speed of PPI prediction	34
2.3.4 Characterizing sensitivity, specificity, and precision	34
2.3.6 Yeast manipulations	36
2.4 Results	35
2.4.1 Conservation of interaction codes	36
2.4.2 Predictions of cross-species PPIs using interaction codes	41
2.4.3 Independent comparison of our approach and the impact of class imbalance	44
2.4.4 Effects of homologous sequence	48
2.4.5 Computational prediction of <i>S. pombe</i> and <i>C. elegans</i> global interactome	49
2.4.6 Prediction of novel protein complexes	53
2.4.7 Investigation of human dsDNA break repair PPI network	55
2.5 Discussion	58
Chapter 3: <i>TPK1</i> influences the efficiency of non-homologous end joining	60
3.1 Abstract	60
3.2 Introduction	60

3.3	Material and methods	63
3.3.1	Yeast strains and plasmids	63
3.3.2	Cell culture, transfection and PRKACB KO construction	64
3.3.3	Protein-protein interaction prediction	65
3.3.4	Plasmid repair assay	65
3.3.5	Drug sensitivity test	66
3.3.6	Synthetic genetic array analysis	66
3.3.7	Chromosomal break assay	67
3.3.8	Protein purification and in-vitro fluorescence kinase assay	67
3.3.9	Chromatin immunoprecipitation analysis	68
3.3.10	NHEJ reporter assay in mammalian cells	68
3.4.1	Identification of <i>TPK1</i> as a candidate involved in double stranded DNA break repair	69
3.4.2	<i>TPK1</i> deletion causes sensitivity to DNA damaging drugs	71
3.4.3	Genetic interaction analysis of <i>TPK1</i>	71
3.4.4	Deletion of <i>TPK1</i> reduces the efficiency of plasmid repair through NHEJ	73
3.4.5	Efficient repair of chromosomal breaks is dependent on <i>TPK1</i>	75
3.4.6	Chromosomal break suppression analysis reveals that <i>NEJ1</i> overexpression can compensate the deletion of <i>TPK1</i>	77
3.4.7	Analysis of double mutant strains suggest that <i>TPK1</i> and <i>NEJ1</i> function in parallel to <i>YKU80</i>	78
3.4.8	Tpk1 phosphorylates Nej1 in vitro	80

3.4.9	Tpk1 is recruited to the site of damage and its deletion reduces the abundance of Nej1 upon DSB induction	80
3.4.10	<i>TPK1</i> role in NHEJ is conserved in mammalian cells	82
3.5	Concluding remarks	84
Chapter 4: Actin related protein 6, Arp6, is involved in repair of DNA double stranded break in <i>Saccharomyces cerevisiae</i>		
		86
4.1	Abstract	86
4.2	Introduction	87
4.3	Material and methods	89
4.3.1	Yeast strains and plasmids	89
4.3.2	Protein-protein interaction prediction	90
4.3.3	Plasmid repair assay	90
4.3.4	Repair fidelity assay	91
4.3.5	Drug sensitivity test	91
4.3.6	Synthetic genetic array analysis	91
4.3.7	Chromosomal repair assay	92
4.3.8	Homologous recombination assay	92
4.4	Results	93
4.4.1	PIPE interactions of ACTL6B homolog of <i>ARP6</i>	93
4.4.2	<i>ARP6</i> mutants reduce efficiency of repair in a plasmid based assay	95
4.4.3	Deletion of <i>ARP6</i> promotes error-free repair by NHEJ	97
4.4.4	Arp6 is involved in repair of chromosomally induced breaks	97

4.4.5	<i>ARP6</i> deletion is sensitive to DNA damaging drugs _____	99
4.4.6	<i>ARP6</i> genetically interacts with several genes involved HR, NHEJ and cell cycle	102
4.4.7	Role of <i>ARP6</i> in NHEJ is related to RSC complex and independent of SWR1 and SWC _____	105
4.4.8	Additional genetic evidence for the association of <i>ARP6</i> with <i>RSC2</i> and <i>MRE11</i> _____	107
4.4.9	<i>ARP6</i> is involved in HR repair _____	99
4.6	Concluding remarks _____	109
Chapter 5: PAS kinase complex regulates the efficiency and fidelity of non-homologous end joining repair _____ 111		
5.1	Abstract _____	111
5.2	Introduction _____	112
5.3	Materials and methods _____	114
5.3.1	Yeast strains and plasmids _____	114
5.3.2	Plasmid repair assay _____	115
5.3.3	Repair fidelity assay _____	115
5.3.4	Drug sensitivity test _____	115
5.3.5	Synthetic genetic array analysis _____	116
5.3.6	Chromosomal repair assay _____	116
5.4	Results _____	117
5.4.1	Plasmid repair analysis suggest a role for <i>PSK1</i> and <i>PSK2</i> in NHEJ _____	117
5.4.2	Deletion of <i>PSK1</i> or <i>PSK2</i> affects the fidelity of DSB repair _____	119

5.4.3	Chromosomal break analysis confirms the influence of <i>PSK1</i> and <i>PSK2</i> on the efficiency of NHEJ	119
5.5.4	Drug sensitivity profiles links <i>PSK1</i> and <i>PSK2</i> to DNA repair	122
5.5.5	Genetic interaction analysis support a role for <i>PSK1</i> and <i>PSK2</i> in DNA repair	124
5.5	Concluding remarks	128
Chapter 6: Conclusion		130
6.1	Final thoughts and concluding remarks	130
6.2	Future prospect	135
References		139
Appendix A		158
Appendix B		159

List of tables

Table 2.1	Summary of information gathered for different organisms tested_____	37
Table 2.2	Definitions of performance measures used in this paper _____	37
Table 2.3	Chosen operating points for the various organisms tested _____	40
Table 2.4	Comparison of this work with previous works_____	47
Table 2.5	Percentages of <i>S. pombe</i> , <i>C. elegans</i> and <i>H. sapiens</i> pairs in which both partners share the same GO SLIM annotation as well as third party interactions_____	51

List of figures

Figure 1.1	DSBs are repaired by two main pathways A) HR and B) NHEJ _____	10
Figure 1.2	Genetic interaction analysis reveal functional redundancy between genes __	21
Figure 2.1	ROC (Receiver Operating Characteristic) curve for <i>C. elegans</i> , <i>E. coli</i> , <i>H. sapiens</i> , <i>S. cerevisiae</i> and <i>S. pombe</i> _____	39
Figure 2.2	Cross-species prediction _____	42
Figure 2.3	ROC curve illustrating the accuracy of cross-species PPI predictions ____	43
Figure 2.4	Accuracy vs. specificity for this work compared to previous work ____	45
Figure 2.5	A novel five member complex for <i>S. pombe</i> identified using interaction codes _____	45
Figure 2.6	Inferring protein function from predicted PPIs _____	57
Figure 3.1	Predicted PPIs and GIs for PRKACB and <i>TPK1</i> _____	70
Figure 3.2	Deletion of <i>TPK1</i> reduces the efficiency of repair for plasmids with overhangs and blunt ends _____	74
Figure 3.3	Deletion of <i>TPK1</i> reduces repair efficiency of chromosomally induced DSB breaks _____	76
Figure 3.4	Kinase activity of Tpk1 and CHIP analysis _____	81
Figure 3.5	PRKACB is critical for NHEJ _____	83
Figure 4.1	PIPE predicted five novel interactions for ACTL6B _____	94
Figure 4.2	Deletion of <i>ARP6</i> reduced efficiency of repair of breaks with cohesive ends _____	96

Figure 4.3	Deletion of <i>ARP6</i> results in enhanced repair of breaks with blunt ends and fidelity of repair _____	98
Figure 4.4	Cells carrying <i>ARP6</i> deletion show high sensitivity to chromosomally induced breaks _____	100
Figure 4.5	Viability of <i>arp6Δ</i> mutants were reduced in under DNA damaging conditions _____	101
Figure 4.6	genetic interaction map of <i>ARP6</i> _____	104
Figure 4.7	deletion of <i>ARP6</i> dramatically reduces repair through HR _____	108
Figure 5.1	Deletion of <i>PSK1</i> and <i>PSK2</i> reduced repair efficiency of breaks with cohesive and non-cohesive ends in a plasmid based assay _____	118
Figure 5.2	Repair accuracy of <i>psk1Δ</i> and <i>psk2Δ</i> mutants strains are reduced compared to WT _____	120
Figure 5.3	Deletion of <i>PSK1</i> and <i>PSK2</i> reduced viability under DSBs induction ____	121
Figure 5.4	Deletion of <i>PSK1</i> and <i>PSK2</i> showed different profile in sensitivity to DNA damaging agents _____	123
Figure 5.5	Genetic interaction maps of <i>PSK1</i> and <i>PSK2</i> _____	127

List of appendices

Appendix A	List of publications_____	158
Appendix B	_____	159
Appendix 2.1	Ordered list of PPIs predictions_____	161
Appendix 2.2	List of human interactions_____	159
Appendix 2.3	Impact of the positive: negative ratio on the precision_____	163
Appendix 2.4	Comparison of precision over human proteins for competing methods_____	164
Appendix 2.5	ROC curve with and without homologous sequences for <i>C. elegans</i> _____	165
Appendix 2.6	ROC curve with and without homologous sequences for <i>E. coli</i> ____	166
Appendix 2.7	ROC curve with and without homologous sequences for human____	167
Appendix 2.8	ROC curve with and without homologous sequences for <i>S. cerevisiae</i> _____	168
Appendix 2.9	ROC curve with and without homologous sequences for <i>S. pombe</i>	169
Appendix 2.10	Amino acid distributions within proteins in different organisms____	170
Appendix 2.11	Difference between using tuned and untuned parameters to predict human PPIs_____	171
Appendix 3.1	Overexpression of key genes involved in NHEJ did not compensate for deletion of TPK1_____	172
Appendix 3.2	Schematic representation of the pEGFP-Pem1-Ad2 reporter_____	172

Appendix 3.3 Multiple sequence alignment of gDNA from empty vector- and sgRNA-mediated PRKACB knockout-transfected U2OS cells____173

List of abbreviations

3-D	3-dimensional
6-4 PP	pyrimidine-pyrimidone photoproducts
AD	Activation domain
Alt-NHEJ	Alternative non-homologous end joining
AMP ^r	Ampicillin resistance
AP	Apurinic or Apyrimidinic
BiFC	Biomolecular fluorescence complementation
CBP	Calmodulin binding site
C-NHEJ	canonical-non-homologous end joining
CPD	Cyclobutane pyrimidine dimers
DBD	DNA binding domain
DDR	DNA damage response
dHJ	Double Holiday junction
D-Loop	Displacement loop
DSB	Double strand break
EGTA	Ethylene glycol tetra-acidic acid

FRET	Förster resonance energy transfer
GAL	Galactose
GI	Genetic interaction
HJ	Holiday junction
HR	Homologous recombination
HU	Hydroxyurea
Leu	Leucine
MMEJ	Micro-homology mediated end joining
MS	Mass spectrometry
Nat ^r	Nourseothricin resistance
NHEJ	Non-homologous end joining
PHLM	Phleomycin
PIPE	Protein-protein interaction prediction engine
PPI	Protein-protein interaction
PSA	Phenotypic suppression analysis
ROS	Reactive oxygen species
SAM	S-adenosylmethionine

SDL	Synthetic dosage lethality
SDSA	Synthetic-dependent strand annealing
SGA	Synthetic genetic analysis
SSB	Single strand break
TAP-tag	Tandem affinity purification tag
TF	Transcription factor
Ura	Uracil
WT	Wild type
Y2H	Yeast 2 hybrid
YPD	Yeast extracts peptone dextrose
YP-GAL	Yeast extracts peptone galactose
ChIP	Chromatin immunoprecipitation

Chapter 1: Introduction

1.1 DNA damage Causes and Types

The genetic information of living organisms is generally encoded within DNA molecules. The coding regions of DNA is transcribed and translated into proteins, which carry out different functions within the cell. Moreover, part of the non-coding DNA sequences play regulatory roles in DNA replication and gene expression processes. These are regions that act as binding sites for proteins involved in replication, gene activation or repression and chromatin remodeling. Both reading and recognition of coding and non-coding sequences are carried out with high specificity and accuracy for effective functionality. It is vital for the cell to keep the integrity of its DNA as alteration to the genetic material could lead to serious consequences for the cell function and survival. DNA is constantly damaged by exogenous and endogenous agents endangering its integrity, identity and stability (Skoneczna A et al. 2015). DNA damage is defined as any alteration to the base, nucleotide or a break in the backbone of the DNA. The main endogenous causes of DNA damage are replication errors and bi-products of metabolic processes and biochemical reactions such as reactive oxygen species (ROS) (Zeman MK and Cimprich KA 2014 and Ciccia A and Elledge SJ 2010). Similarly, DNA could be damaged by exogenous agents including UV radiation, ionizing radiation and environmental toxins (Friedberg EC et al. 2004).

Endogenous factors such as reactive molecules and replication stress are responsible for the majority of damage to DNA (De Bont R and van Labrebeke N 2004). These sources of damage can result in chemical modification of nucleotides and create breaks in the DNA

strands. One of the main sources of endogenous damage is ROS, which are produced as by-product of cellular metabolism. These highly reactive molecules including hydrogen peroxide, hydroxyl radicals and superoxides, can cause over 20 types of base modifications, result in oxidation of nucleotides and induce single or double stranded breaks (De Bont R and van Labrebeke N 2004). Other endogenous DNA damaging molecules are alkylating agents which include S-adenosylmethionine (SAM), a well characterized endogenous methyl donor (B. Sedgwick et al 2007). Endogenous alkylating agents function by addition of a methyl group to nucleotides, mainly adenine and guanine, producing DNA lesions that lead to mutations and block replication machinery (Shrivastav N et al 2010). Other forms of nucleotide damage are caused by spontaneous hydrolysis of bonds between the base and deoxyribose sugar, which consequently leads to formation of abasic sites (apurinic or apyrimidinic (AP) sites) (Boiteux S and Guillet M 2004). Replication stress is another intrinsic source of DNA damage that is triggered by various sources such as mis-incorporation of ribonucleotides, repetitive DNA sequences, nucleotide depletion and unrepaired or mis-repaired DNA lesions (Zeman Mk et al 2014). Chemically modified nucleotides and AP sites block progression of DNA polymerase, triggering replication stress which leads to formation of DNA single stranded breaks (SSBs) or double stranded breaks (DSBs).

Besides intrinsic sources of DNA damage, exogenous agents can compromise DNA integrity and stability. Examples of extrinsic sources of DNA damage include UV and ionizing radiation which form DNA lesions directly and indirectly (Ikehata H and Ono T 2011 and Mehta A and Haber J 2014). DNA damage induced by UV are due to formation of covalent bonds between adjacent pyrimidines, generating lesions in form of cyclobutane pyrimidine dimers (CPD) or pyrimidine-pyrimidone photoproducts (6-4 PP) (Yang W 2011

and Novarina D et al. 2011). On the other hand, ionizing radiation produces base lesions and strand breaks by high energy ionization of the DNA (Close DM et al. 2013). Furthermore, ionizing radiation is able to produce DNA damage through formation of free-radicals such as hydroxyl radicals from water hydrolysis. Direct and indirect sources of ionizing damage contribute to formation of DSBs which are the most severe forms of DNA damage (Mahaney BL et al. 2009).

In case of DSBs formation by stalled replication forks, ionizing radiation, UV, ROS and other factors, cell responses to the break through two main repair pathways, namely non-homologous end joining (NHEJ) and homologous recombination (HR) (Chapman JR et al. 2012). DNA damage response (DDR), which includes several signaling and checkpoint proteins functions upstream of these repair pathways (Jackson SP and Bartek J 2009). DDR plays a central role in cell cycle arrest, histone modification and recruitment of repair proteins to the site of damage (Clemenson C and Marsolier-Kergoat MC 2009). These repair pathways are highly conserved from single cell organisms such as *Saccharomyces cerevisiae* to more complex organisms such as human. The goal of this thesis is to identify several novel *S. cerevisiae* genes that affect NHEJ and to further study them.

1.1.1 DNA double stranded break repair

DSBs are the most detrimental forms of DNA damage caused by endogenous and exogenous agents. They result from breaks created in the phosphodiester backbone of both DNA strands causing DNA fragmentation. DSBs can cause cellular complications including genetic instability, chromosomal aberrations, mutations and cell lethality. These DNA fragmentations compromise genetic stability and integrity due to chromosomal

rearrangement, translocation and deletions. Chromosomal rearrangements and mutations arising from unrepaired and mis-repaired DSBs can lead to activation, suppression or overexpression of genes promoting the onset of important human conditions such as cancer. Moreover, cancer susceptibility and primary immunodeficiency syndrome and neurological disease are associated with mutations in repair genes and defects in the DSBs repair pathways.

Two main repair pathways of DSBs are HR and NHEJ which function in an independent manner. HR repairs DSBs through the use of homologous templates, usually obtained from sister chromatids or homologous chromosomes, to connect the broken sites of DSBs. Since repair is achieved by the use of a homologous sequence, it is considered to be an error-free repair mechanism for DSBs (that is, fewer mistakes) (Ataian Y and Kerbs EJ 2006). This pathway is preferred in case of stalled replication forks and depends on the availability of a homologous sequence (Shrivastav M et al 2008). DSBs can also be ligated directly or by use of micro-homology through NHEJ. This pathway is considered to be error-prone. Recruitment of repair proteins to the site of break is facilitated by a series of signaling proteins and requires histone modification as well as chromatin remodeling (Ataian Y and Kerbs E. J 2006). There are several proteins involved in sensing the damage and preparing the cell for repair through changes in cell cycle progression, regarded as DNA damage checkpoints (Davidson MB and Brown GW 2008). The basis for preference of one pathway over the other remains unclear, however some studies have suggested a role for chromatin remodeling proteins and DNA damage checkpoints in regulation of repair choice (Bao Y and Shen X 2007).

1.1.2 DNA damage response

Once DSBs are induced, a cascade of signaling and sensors will be activated in response to the damage, in DDR pathway. The cell will recognize the damage and in response, it will slow down cell cycle progression or arrest cell cycle to allow enough time to repair the damage (Putman CD et al 2009). DNA damage checkpoints involve proteins that affect cell cycle progression and also trigger initiation of DNA repair. Other than DNA repair proteins, these checkpoint proteins affect key players in transcription, replication, histone modification and cell cycle processes. These proteins are categorized into several groups including damage sensors, adaptor proteins and effectors (Jackson SP 2002). Upon damage, damage sensors will activate signal transducers and pass the information to effectors, which in turn will take the appropriate actions such as delaying cell cycle progression, stalling DNA replication, or recruiting DNA repair proteins to the site of damage (Clemenson C and Marsolier-Kergoat MC 2009).

Core proteins of DDR pathway in yeast are Tel1 and Mec1, which are homologs of human ATM and ATR, respectively (Putman C.D et al. 2009). In case of damage, Ddc2 and RPA will activate Mec1 (Zou L and Elledge SJ 2003) while Mre11/Rad50/Xrs2 (MRX) activates Tel1 independently (Nakada D et al. 2003). There are other DNA damage sensors that are recruited to the site of damage independent of Mec1-Ddc2. Replication factor C (RFC) forms a complex with Rad24, which in turn recruits Rad17-Ddc1-Mec3 complex to the site of damage (Branzei D and Foiani M 2006). The latter complex is responsible for bringing Rad9 and Rad53 to the site of damage and in close proximity of Mec1 and Tel1. Phosphorylation of Rad53 by Mec1 and Tel1 is Rad9-dependant (Zou L and Elledge S.J 2003). Another intermediate that phosphorylates Rad53 is Mrc1 which works independently

of Rad9. Rad53 plays an important role in S/G2 arrest. The other DNA checkpoint effector is Chk1 which is activated by Mec1 in a Rad9-dependant pathway as well. Activation of Chk1 stops cell cycle at metaphase and inhibits the cell from entering anaphase, which shows its importance in G2/M arrest of cell cycle (Branzei D and Foiani M 2006). Delay or arrest in cell cycle provides the cell with enough time to initiate repair of the damaged DNA.

Both kinases, Mec1 and Tel1, phosphorylate H2A to form γ -H2A. Consequently, several chromatin modifying complexes are recruited to the site of damage, and it is suggested that these chromatin modifying complexes might influence the preference of the repair pathway (Attikum HV and Gasser SM 2009). INO80 is one of the complexes, and it has been associated with HR pathway in an MRX dependent manner. INO80 is believed to facilitate recruitment of MRX complex to the site of damage. On the other hand, recruitment of SWR1 complex supports DSBs repair through NHEJ through KU complex (Attikum H et al 2007).

1.1.3 Homologous recombination repair

HR is the primary DSB repair mechanism in *S. cerevisiae*, particularly in DSBs generated by stalled replication forks and during S phase. This error-free mechanism reduces the occurrence of mutation by the use of homologous strand, including sister chromatids, homologous chromosome or repeated DNA sequences, as a reference template for the repair process. Despite HR being considered error-free, mutations could arise due to use of template sequences that are not identical to the break site, resulting in loss of heterozygosity (Shrivastv M et al. 2008). Even though the role of homologous recombination in chromosomal crossover has been known since 1920s, its function in DSBs repair in yeast

was not proposed until early 1980s (Szostak JW et al. 1983, and reviewed in Jasin M and Rothstein P 2015). This discovery triggered investigation of HR role in the mammalian system, identifying the process as a repair mechanism of DSBs through a plasmid based assay (Jasin M et al. 1985 and Jasin M and Rothstein P 2015). HR repair of DSBs is an evolutionary conserved process as the homologs of key proteins involved in HR in yeast are present in mammalian cells; additionally the steps involved in achieving the repair show significance resemblance (Modesti M and Kannar R 2001).

Repair of DSBs by HR is initiated by recognition and resection of DNA break ends, carried out by the MRX complex in yeast (MRN (Mre11, Rad50 and Nbs1) in mammals) and other exonucleases. It is followed by search and invasion of homologous sequence, which is the central step in this process involving Rad51 as the core player (San Filippo J et al. 2008). Subsequently Displacement loop (D-loop) is formed on the homologous template and DNA extension is carried out by DNA polymerases and other factors. The last step of this process involves either the formation of Holiday junction (HJ) which results in crossover and non-crossover product, or displacement of the strand from the D-loop and annealing to non-invading end of DNA through synthesis-dependant strand annealing (SDSA) resulting in non-crossover product (reviewed in Krecji L et al. 2012 and Filippo J et al 2008) (Fig. 1.1A). HR is a complex process requiring a number of proteins to carry out various enzymatic reactions for a successful repair.

The first step in HR is end resection at the site of break of DNA creating 3' overhangs which provides binding sites for other repair proteins. MRX complex in yeast (MRN in human) possess endonuclease and exonuclease activity which is key to DNA end processing at the break site. MRX end resections activity requires Sae2 (CtIP in human), and is limited

to the vicinity of break providing short resection sites (Mimito EP and Symington LS 2008 and Ghodke I and Muniyappa K 2013). Following the initial step, extensive processing of the DNA ends continues either by Exo1 or Sgs1-Dna2 in two parallel mechanisms. Exo1, a highly conserved protein, relies on its 5' to 3' exonuclease activity specific to dsDNA in generating overhangs which could extend up to few kbs (Mimito EP and Symington LS 2008 and Garcia V et al. 2011). Alternatively, Sgs1, ortholog of human BLM, unwinds DNA by its helicase activity, and in conjugation with Dna2 endonuclease processes DNA ends (Zhu H et al 2008). DNA end resection also factors in repair choice regulation, as Cdk phosphorylation of end resection proteins is shown to promote HR and suppress NHEJ (Huertas P et al 2008 and reviewed in Ferretti LP et al 2013). Once the broken ends of DNA are processed, other HR factors will bind to the 3' overhang to complete presynaptic stage.

DNA overhangs created during end resection process are unstable and could form secondary structures. Following end processing, replication protein A (RPA), which has high DNA binding affinity, attaches to the 3' ends of the DNA overhang protecting it from degradation and resolving any possible secondary structures (Sugiyama T et al. 1997). Subsequently, RPA is replaced by Rad51 recombinase in a process assisted by mediator proteins, Rad52 and Rad55-Rad57 complex in yeast and BRCA2 (breast cancer susceptibility 2) in human (Daley J et al. 2013). Rad52 interacts with RPA and RAD51 facilitating displacement of RPA and formation of Rad51 presynaptic filament (Sugiyama T et al. 2009). Besides aiding in Rad51 ssDNA binding, Rad55-Rad57 mediator proteins stabilize Rad51 on the ssDNA and prevent its removal from the presynaptic filaments (Fortin GS and Symington LS 2002). Binding of Rad51 nucleoprotein filament to the dsDNA and search for homologous sequence is facilitated by Rad54, initiating the synapsis stage (Petukhova G et

al. 1998). Rad54 is a motor protein translocase with the ability to transiently break hydrogen bonds between DNA strands as it moves across the DNA duplex. Rad51 filament in conjugation with Rad54 targets the homologous sequence. Furthermore, Rad54 creates D-loop based on its chromatin remodeling activity, and catalyzes strand invasion due to formation of transient strand dissociation (Van Komen S et al. 2000). Following strand invasion, Rad51 is displaced from the ssDNA by Rad54 exposing 3' end for DNA synthesis.

In the final step of the HR, postsynaptic stage, initially DNA polymerases extend the invasive strand preparing it for either direct annealing or formation of double Holiday junction (dHJ). Direct annealing via ssDNA captures the second end of DSB in a process catalyzed by Rad54 and RPA (Sugiyama T et al. 2006). This process which requires D-loop dissolution is suggested to be carried out by Srs2 helicase, promoting non-crossover product (Miura T et al. 2012). Displacement and dissolution of the D-loop results in release of the ssDNA allowing the 3' end to attach to the 5' end of the DSB. Alternatively, the final step may be carried out through a second invasion of the D-loop by the DSB strand forming dHJ. This dHJ could be resolved or dissolved in different ways by set of helicases and resolvases, promoting crossover or non-crossover products, respectively.

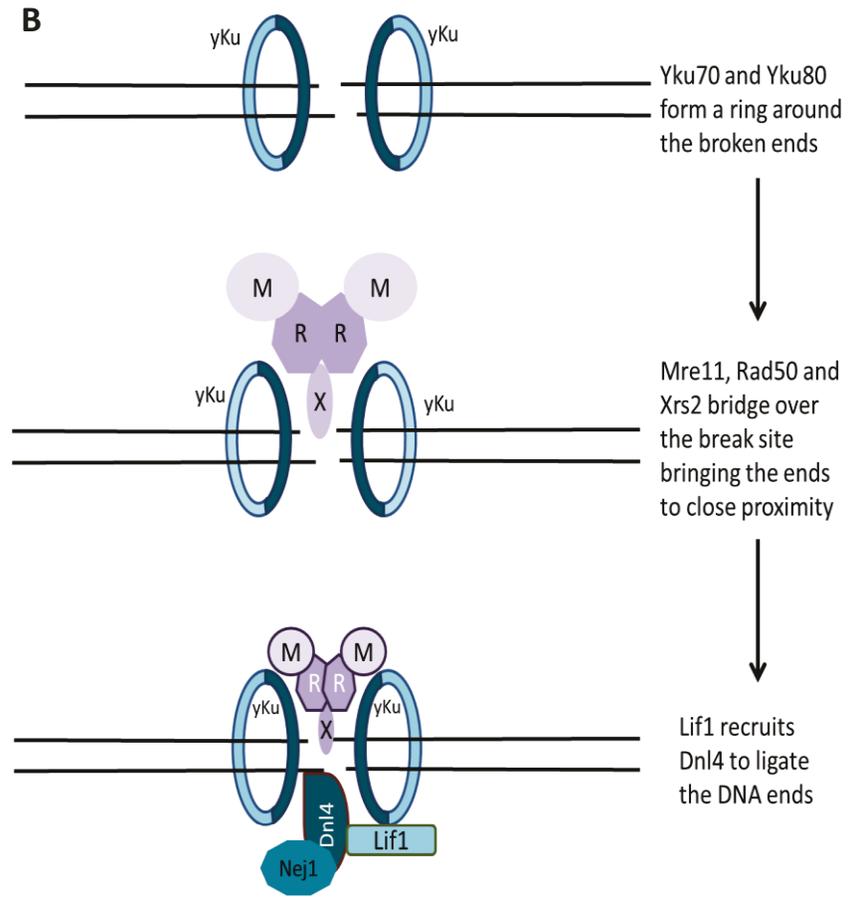
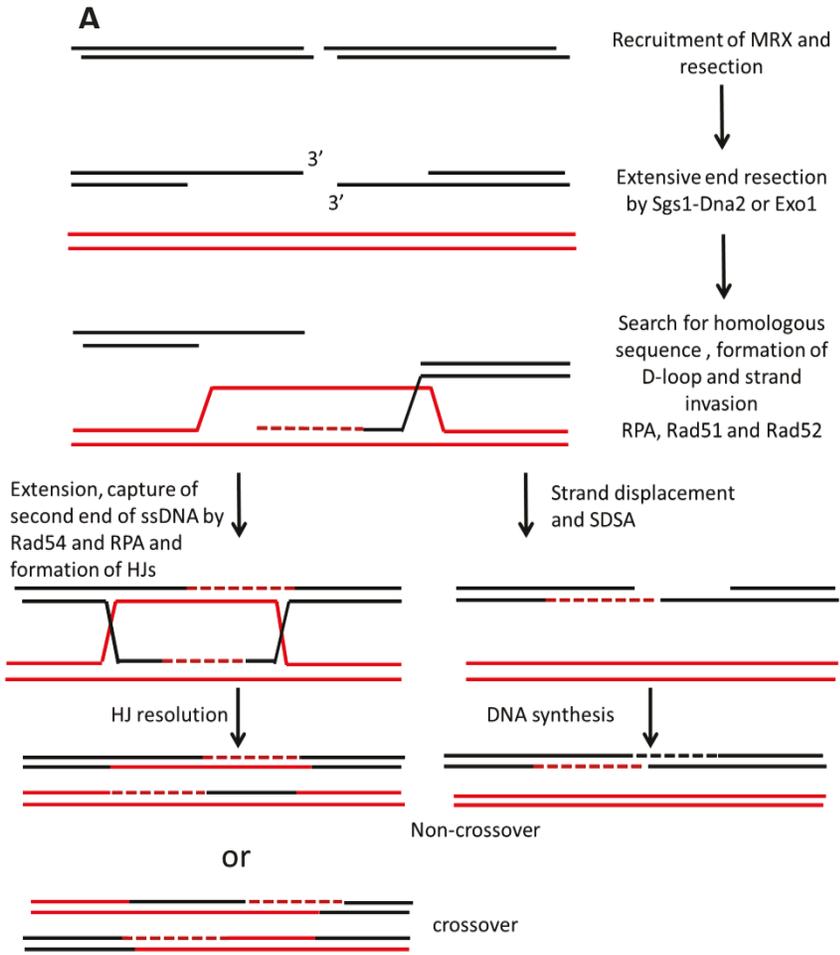


Figure 1.1 DSBs are repaired by two main pathways A) HR which uses a homologous sequence and B) NHEJ which repairs by direct re-joining

1.1.4 Non-homologous end joining

NHEJ repair refers to direct re-joining of the broken ends of DNA using minimum (less than 10 bp) to no homology as a template. It is considered an error-prone pathway, though studies have shown that repair of DSBs with clean overhangs, which do not require end resection, are carried out with high accuracy (reviewed in Bétermier M et al 2014). Mutations resulting from aberrant NHEJ may cause genomic instability which leads to important human conditions such as onset of cancer (Skoneczna A et al 2015). Furthermore, NHEJ is important in immune response development, as its main components function in, V(D)J recombination, a process which promotes variation in receptor genes of immunoglobulin and T-cells (Dai Y et al. 2003). DSBs repair by NHEJ is possible in all phases of the cell cycle, though it has a predominant role in G1 phase due to HR suppression. In mammalian cells, NHEJ is the main repair mechanism of DSBs; however it plays a secondary role in *S. cerevisiae*. The process and mechanism of repair by NHEJ is highly conserved from yeast to mammals; with the exception of DNA-PKCs and Artemis, key NHEJ factors in mammalian cells have orthologues and functional homologs in yeast.

Repair of DSBs by NHEJ is achieved through three major steps: 1) recognition and capture of broken ends of DNA; 2) formation of a bridge to bring broken ends of DNA to close proximity to each other; and 3) ligation and repair of the DSBs (Reviewed by Chiruvella KK et al. 2013 and Shrivastav M et al. 2008 and Daley J et al. 2005). Three main complexes are responsible for each step of the pathway: Yku70/Yku80 for the recognition step; Mre11/Rad50/Xrs2 (MRX complex) is responsible for bridge formation; and Dnl4/Lif1 are involved in ligation and gap filling in the final step (Fig. 1.1B). Besides the key components of NHEJ, other proteins with diverse functions have been associated with this

process. Examples of these factors include: Nej1, functioning in association with Dnl4/lif1; Pol4 and Pol3, playing a role in DNA synthesis and gap filling; and Rad27 carrying out end resection in case of damaged nucleotides. Despite the easy steps, regulation and efficacy of NHEJ as well as its detailed mechanisms appear to be more complicated than originally thought. The list of additional factors that influence NHEJ is continuously growing (Bahmed K et al. 2010, Omid K et al. 2014 and Jessulat M et al. 2015).

First step in NHEJ requires Yku heterodimer, composed of Yku70 and Yku80 (KU70 and KU80 in mammals), the core element of NHEJ pathway. This heterodimer possesses high DNA binding affinity and forms a ring around the broken ends of DNA, stabilizing it and preventing it from degradation. Binding of the Yku heterodimer to the DNA is sequence-independent and only takes place in one direction with the C-terminus of Yku80 facing the DSB end (Daley JM et al. 2005). Once the Yku complex is bound to the break ends, it recruits other factors to the site of break. Mutational analysis has shown importance of Yku80 C-terminus in recruitment of Dnl4 to the site of damage and its anchoring on the break site (Palmbos PL et al. 2008). Although MRX recruitment to the site of damage is independent of Yku *in vivo*, its binding to the site of damage is reduced in the absence of Yku (Wu D et al. 2008 and Zhang Y et al. 2007). Additionally recruitment of Yku and Xrs2 appears to happen at the same time, indicating presence of Yku and MRX complexes at the site of break immediately after damage induction (Wu D. et al 2008).

Second step of repair through NHEJ is carried out by MRX complex, composed of two subunits of Rad50 and Mre11 and one subunit of Xrs2. Even though, MRN, homolog of MRX, is present in mammalian cells, its role in NHEJ is not significant contrary to MRX which is a key component of NHEJ in yeast (Lieber MR 2010). MRX complex forms a

bridge over the site of damage which brings the broken ends closer to each other and prepares them for ligation (Chen L et al. 2001). All members of the MRX complex possess DNA binding affinity; however Xrs2 is the main protein involved in targeting MRX to the site of break (Trujillo M et al. 2003). Rad50, the largest protein of MRX complex, has coil-coil regions that separate its two ATPase domains. These domains are placed on either side of the broken DNA and are connected by an Mre11 dimer, which also mediates interactions between Rad50 and Xrs2 (Hopfner KP et al. 200). Xrs2 also interacts with Lif1 and mediates recruitment of Dnl4/Lif1 to the site of damage; this also requires Yku80-Dnl4 interaction (Palmbos PL et al. 2008 and Matsuzaki K et al. 2008). Furthermore, MRX plays a significant role in the final step by promoting the ligation activity of Dnl4/Lif1 (Chen L et al. 2001).

Final step of NHEJ in yeast is the end-joining of the broken ends which requires Dnl4 and Lif1, functional homologs of human DNA ligase IV and Xrcc4. Lif1 acts as the mediator that interacts with Xrs2 and Dnl4, which performs ligation by forming circular rings around DNA (Chen L et al. 2001). Lif1 recruits Dnl4 to the site of damage in a Yku-dependant manner and promotes its stability and ligase activity at the site of damage (Wu et al. 2008 and Teo S and Jackson SP 2000). Presence of Dnl4 and Lif1 supresses HR by stabilizing Yku complex on the broken ends of DNA and inhibiting DNA end resection and formation of 3' ends (Zhang Y 2007). Another factor suggested to be part of Lif1/Lig4 complex is Nej1, XLF in mammalian cells, which interacts with Lif1, yet its recruitment to the site of damage requires Dnl4 (Wu D et al. 2008 and). Efficient repair of DSBs requires Nej1/Lif1 interaction, as NHEJ repair is impaired by mutations in the interacting domains of Nej1 and Lif1 (Mahaney BL et al. 2014). Additionally, Nej1 catalyzes Dnl4 reactivation, resulting in repeated ligation events by Dnl4 molecules (Chen X and Tomkinson AE 2011).

In addition to the key NHEJ proteins, other factors have been shown to take part in the repair process, and are required for an efficient and accurate repair. Ligation of broken DNA requires clean DNA ends, in case of damaged nucleotides at the DNA termini further DNA processing and resection is needed. For example Pol4, part of Polymerase X family, and Rad27, an endonuclease, function in gap filling and DNA end processing, respectively. Recruitment of Rad27 and Pol4 is dependent on all three members of the ligation complex, Dnl4, Lif1 and Nej1, and also promote their activity in DNA flap removal and DNA synthesis (Tseng H and Tomkinson AE 2004 and Yang H et al 2015). Other factors, such as Exo1 and Tdp1, have been shown to regulate accuracy of repair due to their end processing activity (Bahmed K et al. 2011 and 2010). Additionally, alterations in efficiency of NHEJ has been reported to be linked to deletion of Rtt109, Sub1, Pph3, Psy2 and several other proteins (Jessulat M et al. 2008, Yu L, Volkert MR 2013 and Omidi K et al. 2014). This suggests that despite its simplicity, the pathway of NHEJ requires a wide range of proteins to efficiently catalyze the repair.

Surprisingly, DSBs repair through NHEJ is not completely abolished by deletion of its key factors, suggesting presence of an alternative pathway for repair. Alternative NHEJ (Alt-NHEJ), also referred to as micro-homology mediated end joining (MMEJ), carries out the repair process independent of Yku heterodimer and/or Dnl4, and is characterized by the use of micro-homology. Generally, repair through alt-NHEJ pathway results in deletions and aberrant rejoining leading to chromosomal rearrangement. Mechanism of alt-NHEJ is poorly understood, particularly in yeast, though it is generally activated in the absence of key factors involved in classical or canonical NHEJ (c-NHEJ) (reviewed in Frit P et al. 2014 and Pannunzio NR et al. 2014). In yeast, MRX, Rad1 and Rad10 endonucleases perform DNA

end resection and remove 3' DNA flaps (Lee K and Lee SE 2007). DNA synthesis and extension follows DNA processing which requires a group of polymerases, namely PolH, Rev3/Rev7, Pol4 and Pol32 (Lee K and Lee SE 2007 and Villarreal DD et al. 2012). Other factors such as Nej1, Sae2 and Rad27 have been implicated in alt-NHEJ, though the exact biochemical activity of these proteins in the pathway is yet to be fully understood.

Recent studies have uncovered additional aspects of NHEJ pathway, elucidating the biochemical and molecular mechanism of its key components including their recruitment and localization in response to DNA damage, DNA binding affinity and enzymatic activities. Physical interactions and communications within NHEJ proteins have been extensively studied, revealing regulatory factors within NHEJ. Despite the recent advances in the field, further investigation is required to better understand DSBs repair process, particularly regulation of repair choice, mechanism of Alt-NHEJ and characterization of novel factors which alter repair efficiency/fidelity through NHEJ.

1.2 Systems biology

The principal focus of molecular biology research is examining the role of genes and the structure and function of their products to unravel the pathways, processes and cross-communications in a cellular context. In classical molecular biology studies, generally the focus has been on conducting research on individual genes, and obtaining information and data from independent studies to help in understanding different molecular pathways. These detail oriented studies of individual set of genes did not provide a comprehensive view of the networks and cross-communications within cellular mechanisms and pathways. These limitations combined with the advances in biotechnology and developments of novel

techniques have shifted the field to investigate genes and proteins on a large-scale level in a systematic manner. In recent years development of novel and improved techniques to obtain and analyze large-scale data has been substantially increasing (Schoenrock A et al. 2015, Babu M et al. 2011, Jäger S et al. 2011, Guénolé A et al. 2013 and Hu P et al. 2009). Data generated from these studies can provide insight into functions of individual elements of the cell as well as the cross-communication between various processes. This approach to molecular biology research has been adopted by different disciplines emerging the “-omic” era of research.

Systems biology refers to the investigation of interactions and communications between components of the cell and pathways of a biological system using a comprehensive experimental design (Conesa A and Mortazavi A 2014). In the past two decades with the emergence of fields of genomics and proteomics, and generation of large amount of data, systems biology has gained considerable interest. Systems biology is a hypothesis-driven field which focuses on developing novel techniques and improved experimental designs to explain complex biological systems on molecular basis (Chuang H-Y et al. 2010). It provides insights into functional relationships between genes and pathways and examines cellular behavior in response to perturbation, such as environmental toxin or genetic alterations, using molecular and computational techniques. One of the main goals of system biology is developing computational tools capable of predicting different cellular dynamics and networks, such as gene expression and interaction networks, under various environmental and cellular conditions (Aggarwal K and Lee KH 2003 and Aloy P, Russell RB 2006). Information obtained by these systematic studies can be applied to address questions in

regard, but not limited, to functional annotations, disease implications and network cross-talk.

1.2.1 Functional genomics

In recent years genome-wide sequencing of many eukaryotic and prokaryotic have provided information regarding the location of protein coding genes within the genome. On their own, these sequence information do not provide information regarding the function of the gene and dynamic of the cell, however it is an essential starting point for follow up experiments. Functional genomics refers to investigation of functions of genes and their products using large-scale experimental designs. Data gathered from high-throughput arrays such as genome sequencing, gene expression analysis, genetic interactions (GIs) arrays and protein-protein interaction (PPIs) networks are used to understand the role that genes play in a cell. Due to large amount of data obtained by the large-scale screenings, systems biology methods are generally complemented with computational analysis to obtain meaningful information.

S. cerevisiae has been extensively utilized as a model organism to investigate gene functions in eukaryotic cells (Samanfar et al. 2013, Makhnevych T et al. 2009 and Sharifipoor S et al. 2012). Availability of various resources such as genomic sequence and gene deletion library in addition to ease of genetic manipulation have helped yeast to emerge as an advantageous model organism in systems biology and functional genomics. Approximately two decades ago, development of yeast gene deletion library was initiated to investigate function of uncharacterized genes. Construction of this array has enabled genome-wide investigation of gene functions by examining the relationship between

genotype and phenotype. One application of this array has been in drug sensitivity assays, observing phenotypes of the mutants in the presence of drugs or chemicals. Sensitivity of mutants to a specific drug implies involvement of genes in pathways that are targeted by those drugs (Alamgir M et al. 2010 and Galván Márquez I et al. 2013). This technique has been useful in identification of genes involved in DNA repair, translation, oxygen sensing and cell wall synthesis (Samanfar B et al. 2013, Jessulat M et al. 2015 and Elsztein C et al. 2011). Alternatively deletion arrays can be applied to investigate the mode of activity of various drugs. In this assay, referred to as chemical genomics, gene mutants that are sensitive to a drug of interest are clustered based on their molecular function and/or biological process identifying pathways that are compromised in response to drug induction (Galvan IJ et al. 2008).

Additionally functional genomics in yeast have greatly benefited from genome-wide studies elucidating the expression profiles and GIs and PPIs network maps (Bellay J et al. 2011, Fiedler D et al. 2009 and Ooi SL et al. 2001). Functional analysis using expression profiles have advanced with the development of DNA and protein microarray technologies (Hughes TM et al. 2000, Liu J et al. 2015 and Klutstein M et al. 2010). These studies identify genes presenting similar expression pattern in response to external or enteral stimuli, assigning novel functions to the previously uncharacterized genes. On the other hand, interaction maps identify functionally related genes by examining the role of their interacting partners in the cell. These global genetic and proteomic analyses have produced a large amount of data for better elucidating gene regulation and protein function in various pathways and cellular processes. Despite the advances in functional genomics and

availability of data, there are hundreds of uncharacterized genes in yeast requiring further investigation.

1.2.2 Genetic interaction

GI determines functional dependency and redundancy of genes within cellular pathways (Costanzo M et al. 2011). It investigates the relationship between genotypes and phenotypes in a systematic manner providing complex interactions networks. In a simplified manner, GI is often defined as an unpredictable phenotype resulting from deletion of two genes that differs from the phenotype produced by deletion of each gene individually. GI between two genes occurs when double deletion of the genes results in a phenotype that is unexpected and different from deletion of each gene individually. Alternatively, GI could result from deletion of a single gene combined with overexpression of a different gene producing an unexpected phenotype that cannot be explained by observing the phenotype of the single gene deletion or overexpression (Boone et al. 2007). These forms of interactions are generally divided into two broad categories, positive interactions and negative interactions. Positive interactions refer to enhancement of the cell fitness as a result of combined genetic variation relative to individual genetic variation. On the other hand, combined genetic variations resulting in defective fitness or in the most severe cases, cell lethality, are referred to as negative interactions (Boone C et al. 2007).

S. cerevisiae has served as a powerful model organism to study genetic interactions in eukaryotic cells. Synthetic genetic array (SGA) methodology enables generation of double mutants with detectable phenotypic fitness in *S. cerevisiae* on a genome-wide scale (Tong AH and Boone C 2006). In this method, single deletion of target genes are constructed in α -

mating type and crossed to yeast deletion library in a-mating type. Following a series of selections a progeny carrying double mutants is acquired. Availability of the yeast deletion libraries, in α - and a-mating, enables systematic large-scale investigation of yeast GIs in timely manner.

SGA is a systematic tool which maps interaction networks of the cell identifying functionally related genes and overlapping pathways (Costanzo M et al. 2011). Generally, deletion of two genes (A and Y) involved in parallel pathways results in negative genetic interaction, referred to as synthetic sick or synthetic lethal (Fig. 1.2). In this arrangement, gene Y compensates for the absence of gene A and vice versa in single deletion strains, though in case of double deletion of A and Y, the compensatory pathway is dysfunctional causing defective fitness. Synthetic lethal interactions are of particular interest, as they indicate overlapping pathways that are essential for cell viability. On the other hand deletion of two genes (X and Y) involved in the same pathway may lead to positive interaction, referred to as synthetic suppression (Fig. 1.2). In this situation, single deletion of X or Y, which function in a linear pathway, causes defects in fitness, however double deletion of X and Y completely eliminates the pathway and triggers another pathway which compensates for the absence of both genes (Bryshnikova A et al. 2013).

SGA analysis carried out under different stress conditions are referred to as conditional SGA. This method examines GI network dynamics in response to different stimuli, expanding the interaction map and investigating pathway cross-talk due to perturbation (Bandyopadhyay S et al. 2010). Another modified version of SGA, named synthetic dosage lethality has been developed to study the relationship between overexpression of one gene and deletion of another gene (Sopko R et al. 2006). Additionally,

phenotypic suppression analysis (PSA) studies the effect of gene overexpression on mutant strains under different stress conditions (Alamgir M et al. 2008).

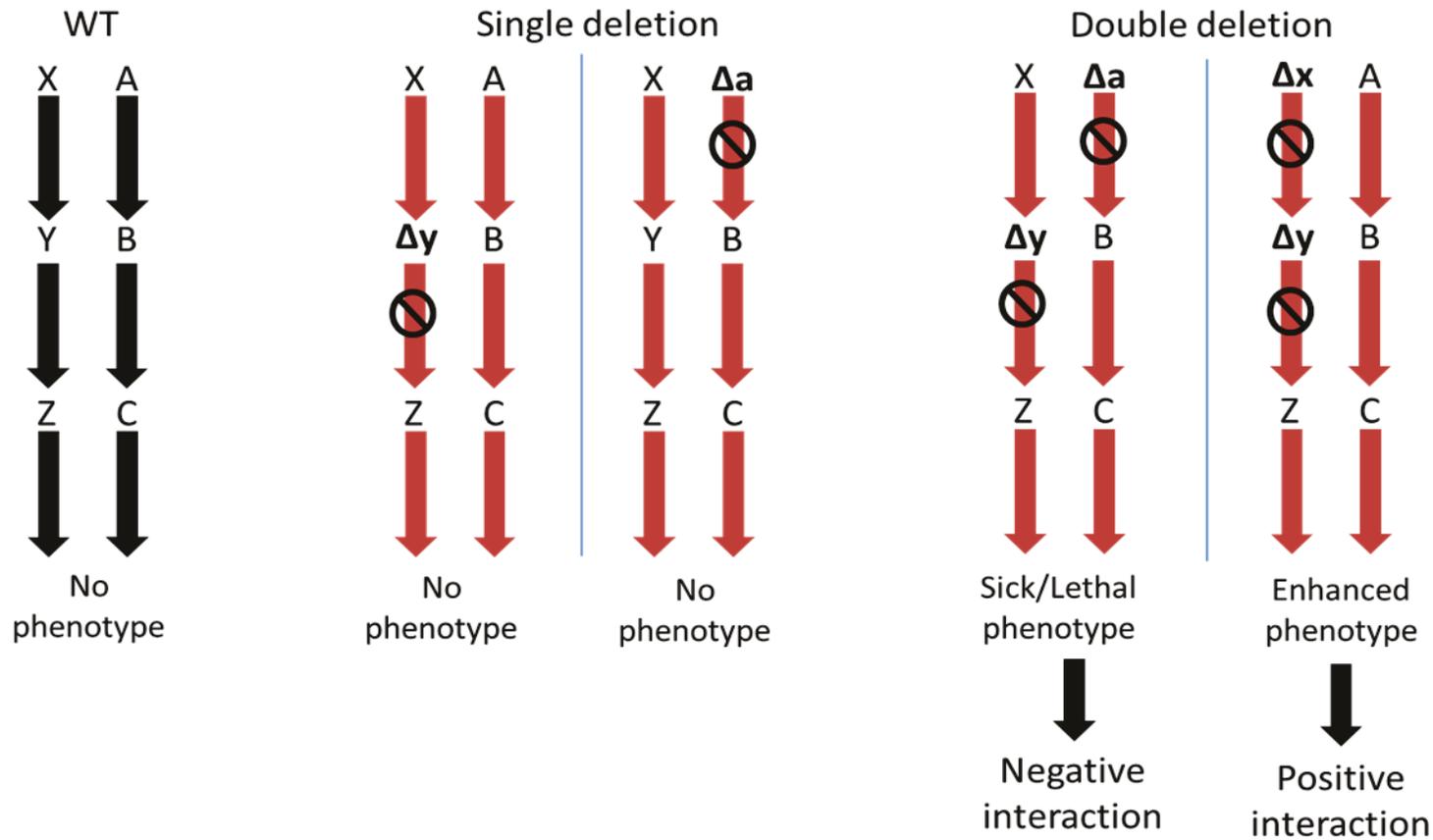


Figure 1.2 Genetic interaction analysis reveal functional redundancy between genes. Deletion of A and Y individually results in no phenotype. Negative interaction refers to sick or lethal phenotype resulting from deletion of two genes functioning in parallel pathways (A and Y). Positive interaction refers to enhanced phenotype resulting from deletion of two genes functioning in the same pathway (X and Y).

1.2.3 Protein-protein interaction

Proteins play an essential role in carrying out cellular functions. Many of these cellular functions, pathways and processes are mediated by proteins that interact with each other to form complexes. It is generally accepted that functionally related proteins interact with each other and it is through these PPIs that proteins realize their functions. PPI is one of means by which internal components of a cell communicate with each other and allow a cell to function as a system (Krogan NJ et al. 2006). Consequently studying PPIs has become one of the main gateways leading to the identification of proteins' functions in biological processes (Pitre S et al. 2008 and Jessulat M et al. 2011). PPIs within a complex can be classified as permanent, in which proteins of the complex have higher affinity for each other and interact longer. These interactors are usually purified as subunits of the same complex. There are also transient interactions which are temporary and often happen under specific cellular and physiological conditions (Perkins JR et al. 2010, Krogan NJ et al. 2006). These interactions, transient and permanent, are guides to mapping biochemical pathways and understanding the mechanism of cellular processes. Recent advances in proteomics have allowed systematic analysis of PPIs networks producing large-scale data and genome-wide interactome of various species.

Over the past decade, several experimental and computational methods have been developed to study PPIs and protein networks within a cell. The two well-established experimental methods for detection and analysis of global PPIs are yeast 2 hybrid (Y2H) and tandem affinity purification (TAP) tag coupled with mass spectroscopy (MS). Generally Y2H is carried out under conditions that do not necessarily represent the natural environment

of the proteins, therefore TAP tagging is advantageous over Y2H method since it is achieved under physiological conditions (Jessulat M et al. 2011 and Brückner A et al. 2009).

Y2H utilizes modularity of eukaryotic transcription factor (TF) proteins to detect PPIs (Fields S and Song O 1989). In this method, TF is divided into its two functional domains: transcription activation domain (AD) that recruits transcription machinery and DNA binding domain (DBD). Subsequently, these domains, AD and DBD, are fused to prey and bait proteins, respectively. Interaction of bait and prey proteins brings AD and DBD in close proximity reconstituting functional TF which in turn activates transcription of a reporter gene (reviewed in Chien CT et al. 1991 and Brückner A et al. 2009). The most common transcription factor domains used in Y2H studies are AD and DBD of GAL4. This TF is linked to the promoter of reporter genes such as *LacZ* which produces b-galactosidase, a quantifiable colored product, or HIS, URA3 and ADE2 auxotrophic selection markers (Suter B et al. 2008). In recent years, development of novel techniques namely Biomolecular fluorescence complementation (BiFC) and Förster Resonance Energy Transfer (FRET), which have adopted Y2H concept, have given the possibility to visualize PPIs through the use of fluorescence markers (Kerppola TK 2006).

TAP-tagging was developed in *S. cerevisiae* to purify and analyze targeted protein complexes under native conditions (Puig O et al. 2001). Over the past two decades, this technique has been applied to plants (Stotz HU et al. 2014), mammalian (Knuesel M et al. 2003) and bacterial cell (Hu P et al. 2009) lines to investigate PPIs *in vivo*. In this method, the protein of interest is fused to a double tag at its C- or N-terminus and isolated in two rounds of affinity purification (Jessulat M et al. 2010). The original construct of tag was composed of calmodulin binding protein (CBP), TEV protease cleavage site and two IgG

binding proteins (Puig O et al. 2001). Once the tagged protein is expressed in the host organism, it undergoes two rounds of washes following cell lysis. In the first round tagged protein and its interacting partners bind to IgG beads through its IgG binding protein. Target protein is then released from the beads by cleavage induced by TEV protease. Subsequently, calmodulin beads are introduced which bind to calmodulin binding peptide on the protein of interest. Finally, ethylene glycol tetra-acidic acid (EGTA) is used to release the protein and its interacting partners from calmodulin beads. Once the tagged-protein is affinity purified along with its interacting partners, the complex is resolved using SDS gel electrophoresis and the constituent of each subunit is identified using MS (Forler D et al. 2003, Puig O et al. 2001 and reviewed in Xu X et al. 2010). TAP-tagging is a powerful method in identifying protein complexes, detecting novel PPIs, studying functional relation between proteins and performing genome-wide large scale analyses (Krogan NJ et al. 2006, Butland G et al. 2005, Nassa G et al. 2011 and Hu P et al. 2009).

Even though a substantial amount of data has been gathered using experimental approaches such as TAP-tagging and Y2H, they all have inherent disadvantages. One of their main disadvantages is that they are time consuming and resource intensive (Skrabanek L et al. 2008 and Brückner A et al. 2009). Additionally these techniques cannot be applied to all types of proteins, with some proteins being difficult to study (Berggård T et al. 2007). Furthermore, the overlap between PPI networks obtained from large-scale studies using experimental methods is very low, suggesting high rates of false positives and false negatives (De Las Rivas J and Fontanillo C 2010). Therefore computational methods have been developed to predict PPIs to complement experimental techniques.

Most computational methods to predict PPIs are based on the data gathered experimentally. These computational tools use various characteristics and features of proteins, such as structure, sequence and evolutionary conservation to generate interaction networks (reviewed Zahiri et al. 2014 and Jessulat M et al. 2011). Some methods predict PPIs based on 3-dimensional (3-D) structure of proteins using “docking” model (Ohue M et al. 2013 and Keskin O et al. 2008). The advantage of this method is that detailed information regarding interacting sites and confirmation of the interacting proteins is obtained. On the other hand, one of the major shortcomings of this method is that it cannot be applied to all proteins due to availability of a limited number of known protein structures.

Other computational tools apply aspects of evolutionary conservation such as phylogenetic profile and gene fusion to predict PPIs (Skrabaneck L et al. 2008 and Muley VY and Ranjan A 2013). Some evolutionary based methods examine presence or absence of genes with similar phylogenetic profiles in different organisms. In these methods it is hypothesized that interacting partners are functionally related, therefore their presence is required to carry out their function across different species (Marcotte EM et al. 1999). Two major limitations of this method include its inability to predict most essential proteins that are common to all organisms, and its requirement for fully sequenced genomes to operate. Another approach bases PPIs predictions on gene fusion which combines comparative genomics and evolutionary information. The basic idea of this method relies on the fact that fusion of two genes present in one organism may have produced a single multi-domain protein in another organism (Chia J-M and Kolatkar PR 2004 and Liu ZP and Chen L 2012). The fusion genes are predicted to be interacting partners. The main disadvantage of this

approach is the limited number of proteins that have evolved from gene fusion, restricting it to only a subset of proteins.

Several methods have been developed which use primary protein structures to predict novel PPIs (Gomez SM et al. 2003, Pitre S et al. 2008, You Z-H et al. 2014 and reviewed in Park Y 2009). These methods are advantageous over approaches that use structural information since the amino acids sequences of a large number of proteins is available and easily accessible. Protein-Protein Interaction Prediction Engine (PIPE) is a novel approach which uses the primary protein structure to predict an interaction (Pitre S et al. 2006 and 2008). The hypothesis behind this approach is that specific short amino acid regions, that neither resemble domains nor are structure specific, can be responsible for a subset of PPIs (Pitre et al. 2006). These regions consist of short re-occurring sequences that exist in many proteins. According to this approach the presence of co-occurring polypeptide regions in the database of interacting proteins can be used as a basis to detect novel interactions in target proteins. To date, this method has been tested to map the interaction network of *S. cerevisiae*, *Schizosaccharomyces Pombe*, *Caenorhabditis elegans* and human (Pitre S et al. 2006, 2008, 2012 and Schoenrock A 2014).

1.3 Focus and objectives

DSBs are the most severe form of cytotoxic DNA lesions in the cell, constantly endangering genetic integrity, genomic stability and cell viability. These forms of damage are of significant interest due to the difficulty of their repair. Aberrant repair of DSBs increase the chance of chromosomal rearrangement, mutation and apoptosis. DNA termini that are physically separated as a result of DSBs could recombine with different parts of the genome

leading to gene translocation and promoting tumorigenesis. These recombination events could lead to overexpression of oncogenes or deletion of tumor suppression genes resulting in development of various types of cancer cells (Helleday T et al. 2014). Interestingly, DSBs repair pathways are compromised in cancer cells, making DSBs induction a target for cancer therapeutic drugs. Genotoxic drugs, such as bleomycin and hydroxyurea, generate DSBs in cancer cells; impairment of repair pathways in these cells leads to apoptosis (Jekimovs C et al. 2014). Additionally, mutations in genes involved in repair pathways of DSBs, NHEJ and HR, are associated with cancer susceptibility, signifying importance of DSB repair in understanding cancer and therapeutics (Sobanski T et al. 2015 and Deitlein F et al. 2014). Consequently, it is important to investigate details of DSB repair because not only is this a fundamental process within a cell, but also increased knowledge of DSB repair can lead to identification of new targets for development of new therapeutics.

NHEJ is the principal repair pathway of DSBs in mammalian cells and is present throughout the cell cycle phases. This repair mechanism which uses direct end joining, plays a central role in developing a variety of T-cell receptors and immunoglobulin genes via V(D)J recombination processes. Clinical studies have shown patients with immunodeficiency carry mutations in key NHEJ genes, such as XLF, LIG4 and Artemis (Woodbine L et al. 2014). NHEJ is important in repair of physiologically induced DSBs such V(D)J and DSBs resulting from intrinsic and extrinsic agents. Even though the general mechanism of this highly conserved process is known, several new proteins have been associated with the pathway, which alter its efficiency and/or accuracy of repair. The list of these proteins continues to grow (Heo J et al. 2015, Ruiz JF et al. 2013, Jessulat M et al. 2015 and Omid K et al. 2014). This implies existence of other novel proteins that are potentially involved in the

repair of DSBs through NHEJ. This gap of knowledge in the repair pathway requires further investigations. One of the objectives of this thesis is to identify new factors that affect the efficiency of NHEJ in the yeast *S. cerevisiae*.

In functional genomics, elucidating interaction maps, such as PPI maps, is often used to identify functionally related proteins. Therefore, it is expected that identification of novel proteins involved NHEJ could be possible by examining a broad range of PPIs for repair proteins. In this way identification of a novel interacting partner with NHEJ proteins can indicate a new role in the repair pathway. In this thesis we utilize PIPE to generate a computationally predicted local PPI map for DSB repair proteins. In doing so, we identify new proteins that have novel interactions with established DSB repair proteins. We then subject the newly identified proteins to focused follow-up investigations to better understand the mechanism by which these factors influence NHEJ in yeast. Specific objectives of this thesis are as follows:

- 1- To screen for novel NHEJ proteins through prediction of novel PPIs of DNA repair and nuclear proteins in human cells based on primary sequence of known interacting proteins
- 2- To experimentally validate involvement of yeast homologs of the novel candidates in NHEJ repair pathway
- 3- To investigate the role of Tpk1, Arp6 and Psk1/Psk2 in NHEJ repair using biochemical and genetic assays

Chapter 2: Short co-occurring polypeptide regions can predict global protein interaction maps

2.1 Abstract

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).

2.2 Introduction

Protein-protein interactions (PPIs) represent an essential aspect of all biological pathways and signaling mechanisms within a cell, and are reliable indicators of functional associations between proteins. Consequently, a major goal of the post-genomics era has been to elucidate a detailed global map of PPIs within a cell. To date, high throughput attempts, which are both time and resource demanding, have been utilized to study the global PPI networks of only a few model organisms. Computational methods provide an attractive alternative. However, due to computational limitations, elucidating the PPI networks of complex organisms such as human has not been possible. Furthermore, with the exception of

S. cerevisiae (Pitre S et al. 2008a), other organisms are yet to be studied at an all-to-all level where all proteins are analyzed for their abilities to interact with all other proteins, again, due to the computational complexity of most approaches.

Several computational methods for predicting PPIs require in depth knowledge regarding the proteins including structure, sub-cellular location, function, interacting domains, etc (Pitre S et al. 2008b). The obvious drawback of such methods is their limited applicability for predicting interactions in organisms which have limited information available. This is apparent by the use of only model organisms (*e.g. S. cerevisiae*) or widely studied organisms (*e.g. human*) used in the testing and application of such methods.

Some PPI prediction methods, however, are based on sequence data only (Zaki N et al. 2009, Gou Y et al. 2010 and Martin S et al. 2005). A growing body of evidence supports the usefulness of short co-occurring polypeptide sequences (interaction codes) in predicting PPIs in yeast (Pitre et al 2006 and 2008 and Neduva V et al. 2005). As discussed in (Pitre S et al. 2006 and 2008 and Neduva V et al. 2005), there appear to exist a finite number of interaction codes of length around 20 amino acids that mediate a subset of PPIs. However, there is no information on the global applicability of this approach, as it has been until now computationally infeasible to apply these techniques to more complex organisms with larger proteomes. Establishing the conservation of co-occurring polypeptide codes will not only provide further evidence for their activities in mediating PPIs but, more importantly, it would highlight the applicability of this approach to detect genome-wide PPIs in other organisms such as humans. In addition, it allows for the prediction of PPIs in newly sequenced organisms for which limited or no experimental PPI data is available.

Through significant computational acceleration of our previous approach (Pitre S et al. 2006 and 2008) via massive parallelization of our previous software, it is now possible to scan larger and more complex proteomes. In this paper, we will investigate the possibility of predicting PPIs based on sequence data only in various organisms including *S. pombe*, *C. elegans*, *E. coli*, and most importantly human. In addition to our previously published genome-wide scan of the *S. cerevisiae* interactome, we will present two new genome-wide scans (*S. pombe* and *C. elegans*) along with validation and analysis of the resulting predicted interactions. We will also explore part of the human interactome.

2.3 Methods

2.3.1 PPI prediction via interaction codes

In brief, to predict whether two query proteins are likely to interact, our method examines sliding windows of primary sequences to determine if both query proteins share similarity with pairs of proteins that have been previously reported to interact. For more details, please see (Pitre S et al. 2008 and 2006). To measure sequence similarity, the PAM120 substitution matrix is used. The method requires a similarity threshold applied to the PAM120 score to determine whether sequence windows are, in fact, similar. This threshold must be tuned for each organism, as described below. Once we have completed this analysis over all possible sequence windows in each query proteins, a decision must be made whether there is sufficient evidence to support the predicted interaction. A second score threshold is then applied to make the decision whether the interaction should be predicted or not. As discussed below, this second threshold is tunable to achieve the required level of specificity or precision.

2.3.2 PPI prediction adaptation algorithm for other organisms

Amino acid distributions within proteins differ significantly between different organisms; see appendix 2.10. Therefore, it is not optimal to run the same PPI prediction method for different organisms. Appendix 11 shows the difference between simply applying the unmodified PPI prediction method using yeast parameters to human proteins and the adapted human PPI prediction method using the organism-specific PPI prediction adaptation algorithm described in the remainder of this section.

Since amino acid sequence window comparison operations in our software are based on the PAM120 substitution matrix, a different amino acid distribution will cause a shift in the expected window score when two random sequences are compared. We plot the probability of scores when comparing two fragments of length 20 and set our threshold such that two sequence windows are declared to be ‘similar’ only if their PAM120 score is significantly above that expected by chance (i.e. there remains only a probability of 10^{-6} of obtaining this score by comparing random fragments). We have previously determined (Pitres S et al. 2006) this cutoff to be 35 in *S. cerevisiae* and that remains a valid cutoff for most of the other organisms tested except one: *H. sapiens*. For *H. sapiens* the expected score when comparing fragments increased, and a new cutoff of 40 was used exclusively for this organism. Changing the fragment window length (20 AA) and substitution matrix did not offer any significant improvement and they were therefore not modified.

For any new organism, the steps for automatically tuning the window matching score used by our PPI prediction method, precomputation of similar windows, evaluating the method’s performance and selecting an operating point are as follows:

Step 1: Window matching score tuning

- (a) Part of the input is all the protein sequences for the organism (usually from an online source). We can limit the proteins to those involved in known interactions; however the entire proteome is preferred.
- (b) Our method then calculates the amino acid distribution for the organism. This is used to evaluate the expected score for comparing two random fragments of length 20 using this distribution. A score is picked to represent a 10^{-6} chance that two windows of length 20 are matched at random.

Step 2: Precomputation of similar windows

- (a) Once the correct window matching score setting have been determined, the database pre-computation is run. That is, the similarity between each unique 20 AA window in the proteome is precomputed and cached.

Step 3: Performance evaluation and picking an operating point

- (a) The second part of the input is a set of known interactions for the organism. This can be from one experiment, or from a repository such as BioGRID (Van Zuylen WJ et al. 2012). It can also be a gold-standard set or the union of all the databases online for example. Only physical interactions should be used since this is what the algorithm is meant to predict.
- (b) The method then needs to evaluate the sensitivity and specificity of the predictions for this organism. This is done by LOO cross-validation using the known interaction set collected earlier. The negative set is generated from 100,000 random pairs (Ben-Hur

A and Noble WS 2006). Plotting the ROC and recall vs. precision curves for the LOO cross validation experiment will indicate what cutoff to use in order to achieve a given specificity, precision, or sensitivity. For high-throughput experiments, operating at a high specificity is strongly suggested in order to avoid a large number of expected false-positives and maximize the precision of the classifier.

Step 4: PPI predictions

- (a) After completing the automatic steps listed above, the method is ready to start predicting new PPIs for that organism. For small genomes this can be done by an all-to-all experiment but for large genomes it is often preferable to only run pairs of interest due to the runtime involved.

2.3.3 Improved speed of PPI prediction using parallel multi-core processing

An ongoing concern with PPI prediction is performance. For example, for a single human protein pair such as O43345 and Q05481, sequential PPI prediction took 1894.04 seconds on a single processor core. We therefore used fine-grained parallelism to accelerate our method using the Intel Cilk Plus library. For the same protein pair, our new, parallel multi-core PPI prediction method took only 321.82 seconds on a quad core Core i7 860 processor. This represents a 5.9 fold speed improvement. This super linear speedup (more than number of cores) can be attributed to memory latency hiding caused by Intel Hyper-Threading.

2.3.4 Characterizing sensitivity, specificity, and precision of PPI predictions via in silico experiments

To determine the sensitivity and specificity of our PPI predictions we conducted LOO cross-validation experiments by first obtaining interactions databases (Table 2.1) to use as positive

sets. For each individual organism, a negative set composed of 100,000 randomly chosen (Ben-Hur A and Noble WS 2006) pairs was created. The LOO experiments were conducted as per Step 5 of the “Sensitivity and Specificity Measure for Other Organisms” section discussed previously. The results are plotted as a ROC curve representing sensitivity against 1-specificity (Fig. 2.1-2.3). While the ROC curve displays the entire range of specificity (0-100%), high-specificity values are of greater interest due to the relatively low number of true-positives expected for PPI maps, which can easily be outnumbered by false-positives unless the method operates at high specificity. Prediction confidence can also be measured using precision, which measures the proportion of positive predictions which are likely to be true positive interactions. The problem with precision is that it depends directly on the actual ratio of positive-to-negative protein pairs, and this is typically unknown. Previous studies have fixed this ratio at very low values (e.g. Yu CY et al. explored ratios as high as 1:15, but admit that true ratios are likely on the order of 1:1000 for some species). For any given ratio Positive(P):Negative(N), the precision can be directly computed from sensitivity and specificity (which are unaffected by this ratio) using the following equation:

$$\text{Precision (Prec.)} = (\text{Sens.} \cdot \alpha) / ((\text{Sens.} \cdot \alpha) + (1 - \text{Spec.}) \cdot (1 - \alpha)) \quad \text{where } \alpha = P / (N + P)$$

The precision of our method for each species is illustrated for a range of ratio values in appendix 2.3.

2.3.5 Sensitivity and specificity measure for cross-species predictions

To determine if PPIs in a target organism can be predicted from cross-species PPIs using the union of multiple species, all known interactions from our select organisms (*C. elegans*, *E.*

coli, *H. sapiens*, *C. cerevisiae* and *S. pombe*) are used except the interaction from the organism used for testing. For example to test our prediction in *H. sapiens* we would use all known interactions for the organisms above except interactions for *H. sapiens*. This simulates predictions in a new organism or for one which has few known interactions.

2.3.6 Yeast manipulations

The collection of yeast gene knockouts is described in (Winzeler EA et al. 1999) Plasmid repair analysis was performed as before (Jessulat M et al. 2008).

2.4 Results

2.4.1 Conservation of interaction codes

Having previously demonstrated the utility of interaction codes for the prediction of PPIs in yeast (Pitre S et al. 2006 and 2008), here we explore the conservation of interaction codes by investigating whether PPIs in organisms other than yeast may also be predicted from their primary sequences. In the current study, we use co-occurring short polypeptide regions identified from previously reported PPIs to predict novel PPIs. Interaction codes were computationally identified from the available PPI data (see Table 2.1) in four model organisms (*E. coli*, *S. pombe*, *C. elegans*, *S. cerevisiae*) and in humans. Our massively parallel, accelerated software, running on a large-scale parallel computer, is the sole reported experimental technique that is capable of exhaustively scanning all potential PPIs in complex organisms such as *C. elegans* (approximately 23,000 proteins) whose potential interactome contains 2.8×10^8 protein pairs. Unlike other reported techniques that rely on a set of known domains or structures (Pitre et al. 2006), our approach is capable of predicting interactions between completely uncharacterized proteins based solely on primary sequence (and a starting PPI dataset). The performance measures used in the paper are described in Table 2.2.

Table 2.1: Summary of information gathered for different organisms tested (number of protein sequences, number of interactions and the source of the interactions).

Organism	Number of Proteins	Number of Interactions	Interaction Database
<i>C. elegans</i>	23,684	6,607	BioGRID [15]
<i>E. coli</i>	4,290	16,235	EciD [19]
<i>H. sapiens</i>	22,513	41,678	HPRD [20] and BioGRID [15]
<i>S. cerevisiae</i>	6,716	43,591	BioGRID [15]
<i>S. pombe</i>	5,024	2,951	BioGRID [15]

Table 2.2 Definitions of performance measures used in this paper. TP = True positives, FP = False positives, TN = True negatives and FN = false negatives.

Measure (abbr.)	Equation
Sensitivity (Sens.)	$TP/(TP+FN)$
Specificity (Spec.)	$TN/(TN+FP)$
Precision (Prec.)	$TP/(TP+FP)$
Accuracy (Acc.)	$(TP+TN)/(TP+TN+FP+FN)$
F-measure (Fm)	$(2 \cdot \text{Prec.} \cdot \text{Sens.}) / (\text{Prec.} + \text{Sens.}) = 2 \cdot TP / (2 \cdot TP + FP + FN)$

We conducted leave-one-out (LOO) *in silico* experiments to characterize the sensitivity (i.e. relative size of the subset of the true interactions that can be detected by our method) and specificity (i.e. $1 - \text{false positive rate}$) of our approach for each organism (Fig. 2.1) as well as the impact of the positive: negative ratio on the precision (i.e. the proportion of predicted PPIs that will represent true verifiable interactions) (see appendix 2.3 and discussion below). The results in Figure 2.1 are reported using receiver operator characteristic (ROC) curves, which plot the achievable sensitivity at a given false positive rate. Improved performance is reflected in curves with a stronger bend towards the upper-left corner of the ROC graph (i.e. high sensitivity is achieved with a low false positive rate), while a random decision rule results in the diagonal line shown in Figure 2.1. The ROC curves in Figure 2.1 show that our approach can successfully detect PPIs, from the interaction codes, in all five organisms that we investigated. It is critical to operate at extremely high specificities when applying such a method to a genome-wide analysis due to the expected sparsity of the true interactome. Otherwise, in a genome-wide analysis, true predicted PPIs will be vastly outnumbered by false positives. Therefore, insets are provided for each ROC curve highlighting the expected prediction accuracy at very high specificities (99.9 - 99.95%). From Figure 2.1, it appears that our method has the highest accuracy for human proteins, followed by *S. pombe*, *C. elegans*, and *S. cerevisiae*. For example, at a specificity of 99.95%, 23.8% of human PPIs can be predicted. When operating at lower specificities, the accuracy for *S. pombe* surpasses that for humans. The sensitivity for *E. coli* at 99.95% specificity is zero (see inset of Fig. 2.1), however sensitivity surpasses 60% when operating at 80% specificity which is typically useful when analyzing a smaller set of preselected protein pairs. Table 2.3 lists the specificity and sensitivity of our method for each

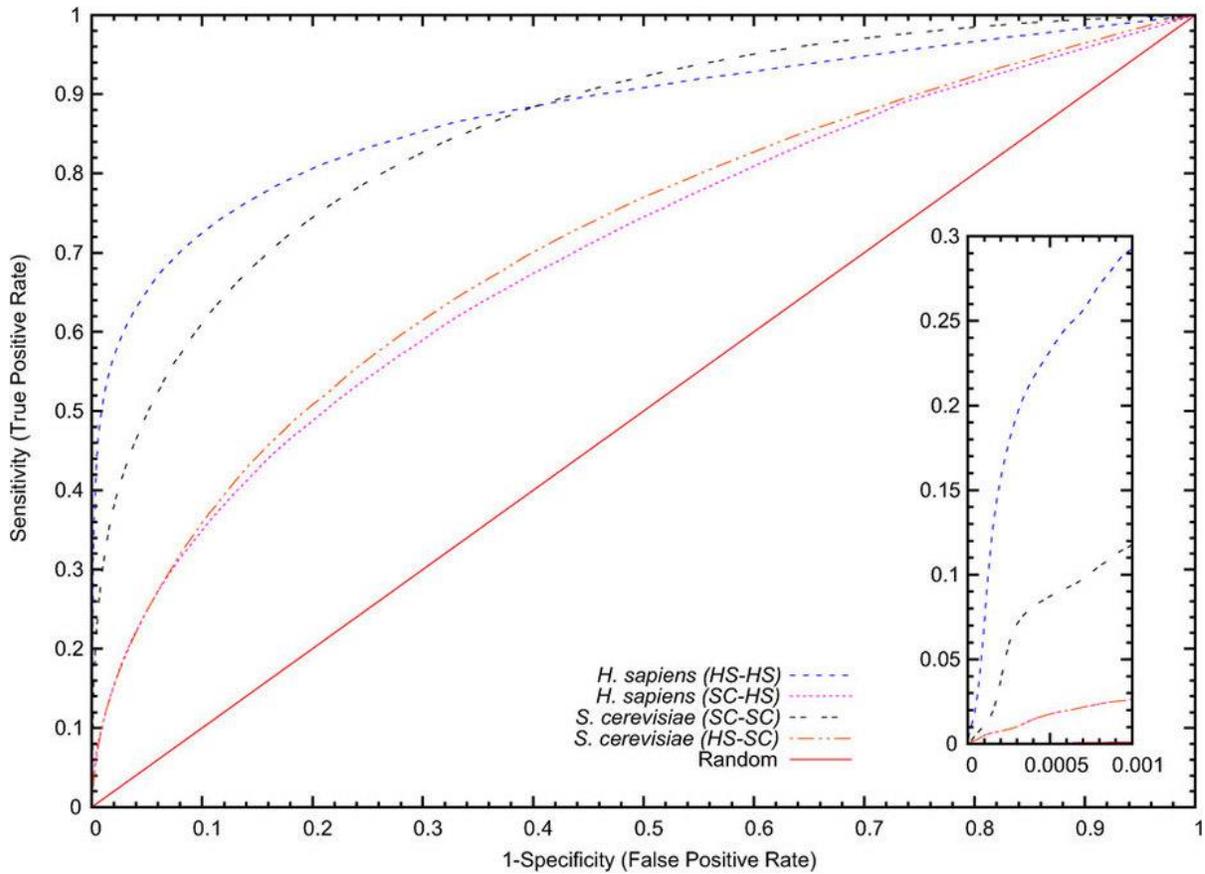


Figure 2.1: ROC (Receiver Operating Characteristic) curve for *C. elegans*, *E. coli*, *H. sapiens*, *S. cerevisiae* and *S. pombe*. The curve presents the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). Inset: performance at very high specificity (99.95%). *H. sapiens* has greater sensitivity than all other organisms tested. Note that due to the scaling of the axes, the diagonal random curve appears flat in the inset.

Table 2.3: Chosen operating points for the various organisms tested. In order to reduce the number of false positives, a specificity of 99.95% is typically chosen (* for *E. coli* a lower specificity of 99.0% was chosen due to the small size of the known interaction set).

Organism	Spec. (%)	Sens. (%)
<i>C. elegans</i>	99.95	9.97
<i>E. coli</i>	99.0*	14.48
<i>H. sapiens</i>	99.95	23.22
<i>S. cerevisiae</i>	99.95	8.77
<i>S. pombe</i>	99.95	16.89

organism at the threshold value used for all subsequent analysis. The impact of homologous sequences in our interaction databases had an inconsistent effect on performance as discussed in detail below.

2.4.2 Prediction of cross-species PPIs using interaction codes

Next we investigated whether PPIs in a target organism can be predicted from cross-species PPIs (Fig. 2.2). Specifically, can we predict PPIs in one organism using known PPIs from a different organism? Note that Park (Park Y 2009) also attempted to answer this question using our previous method (Pitre S. et al. 2008) in conjunction with others. However, (Park Y 2009) used the parameter settings for *S. cerevisiae* in all of his experiments which leads to reduced performance (see discussion below). To answer the above question, we first investigated *H. sapiens* and *S. cerevisiae* because they have the highest number of previously reported interactions. As indicated in Figure 2.2, using known *S. cerevisiae* PPIs to predict *H. sapiens* interactions (SC-HS) is a weaker predictor than using known *H. sapiens* PPIs to predict *H. sapiens* interactions (HS-HS). However, it is surprising that the cross-species predictors (i.e. SC-HS and HS-SC) can still predict meaningful interactions at 80% specificity (46-48% sensitivity) for both organisms. This illustrates that known PPIs in one organism may be successfully used to predict novel interactions in another.

We then investigated the applicability of this approach towards predicting PPIs in one organism from a collection of independent PPIs from an ensemble of several other organisms. For all eukaryotes tested, our ROC curves (Fig. 2.3) suggest again the ability of interaction codes from different species to predict interactions in an independent species. The efficiencies of these predictions are lower than the experiments in Figure 2.1 but similar to

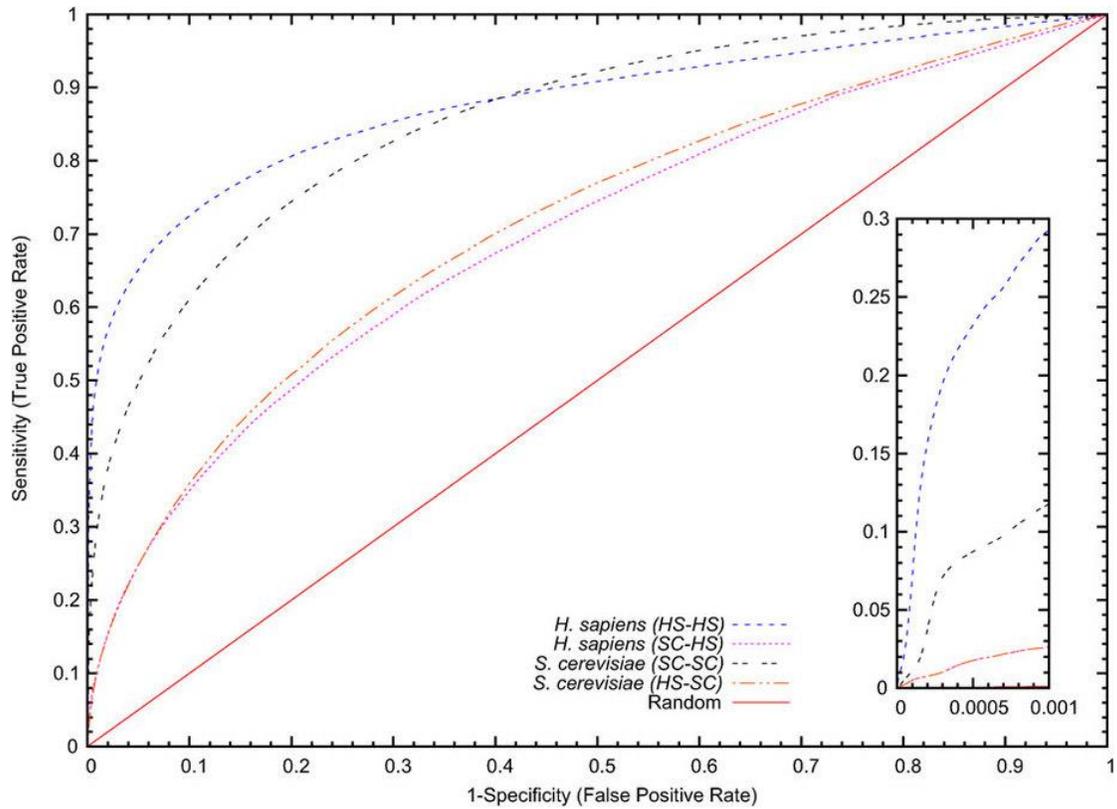


Figure 2.2: Cross-species prediction. Here we see the results of using known *H. sapiens* interactions to predict *S. cerevisiae* interactions (HS-SC) and known *S. cerevisiae* interactions to predict *H. sapiens* interactions (SC-HS) compared to same-species prediction (HS-HS and SC-SC).

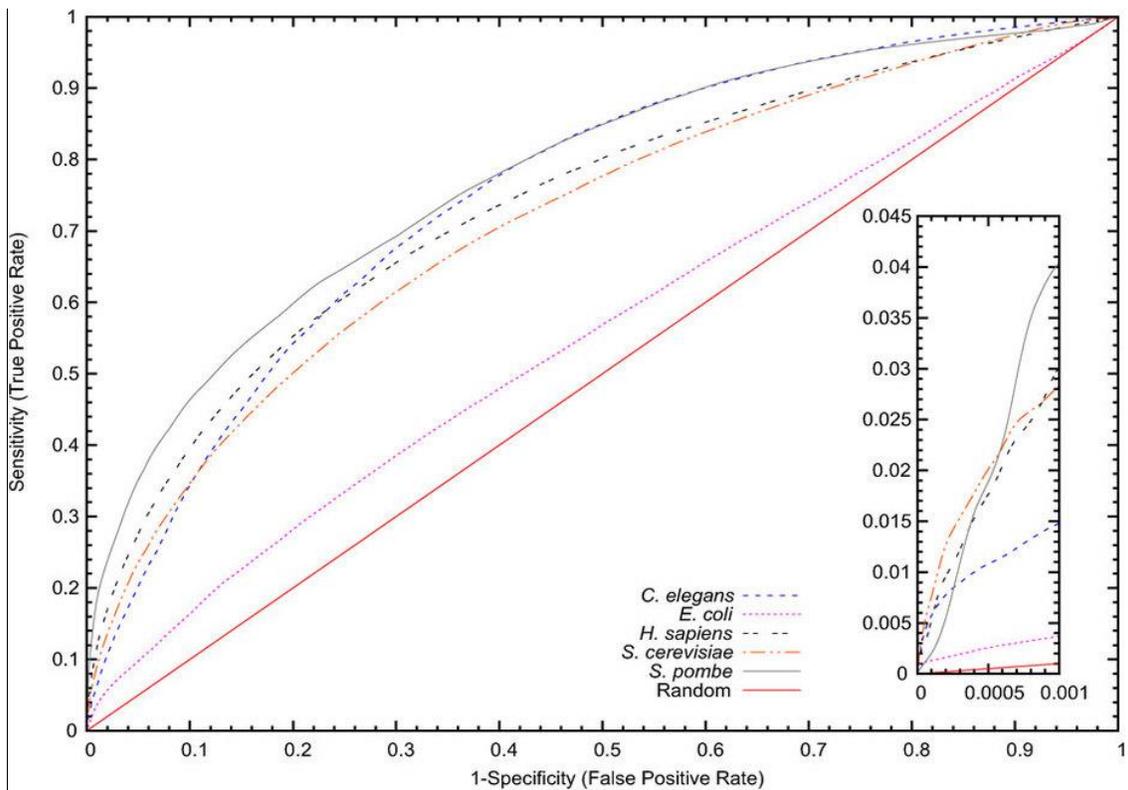


Figure 2.3: ROC curve illustrating the accuracy of cross-species PPI predictions. Here, predictions are based on the union of known interactions from several organisms excluding the test organism. Inset: performance at very high specificity (99.95%). Note that due to the scaling of the axes, the diagonal random curve appears flat in the inset.

the results in Figure 2.2 for *H. sapiens* and *S. cerevisiae*. In fact for *H. sapiens* we see a slight sensitivity improvement at 80% specificity (54-55% sensitivity as opposed to 48-49%) when predictions are based on known PPIs from multiple organisms rather than just *S. cerevisiae*. PPI prediction in *E. coli* seems to be an exception, suggesting differences between conservation of protein codes in *E. coli*, a prokaryote, and the eukaryotic organisms examined here.

2.4.3 Independent comparison of our approach and the impact of class imbalance

Our method has been independently evaluated by Park in (Park Y 2009) and compared against three other approaches (Guo Y et al. 2010, Shen J et al. 2007 and Martin S et al. 2005). Park compared the four methods using four test categories (Park Y 2009): (a) *H. sapiens* data to predict *H. sapiens* PPIs, (b) *S. cerevisiae* data to predict *S. cerevisiae* PPIs, (c) *H. sapiens* data to predict *S. cerevisiae* PPIs and finally (d) *S. cerevisiae* to predict *H. sapiens* PPIs. Tests (a) and (b) evaluate how well each method predicts PPIs in an organism given data about that same organism, while (c) and (d) assess cross-species predictions. Since we already explained the need for high specificities for genome-wide predictions, Figure 2.4 presents the sensitivity results of the experiments in (Park Y 2009) for specificities higher than 95%. We can see that, apart from a close tie with (Martin S 2006) in Figure 2.4A at 95% specificity, our method achieves the highest sensitivity across all four experiments (Fig. 2.4A, B, C and D) at high specificities. In fact at 99.95% specificity, our method attains 3.6 to 17.2 times the sensitivity of the other methods. Also note that Park used the parameter settings for *S. cerevisiae* in all of his experiments even to predict *H. sapiens* PPIs. Amino acid distributions within proteins differ significantly between different organisms; see appendix 10. Therefore, it is not optimal to run the same PPI prediction method

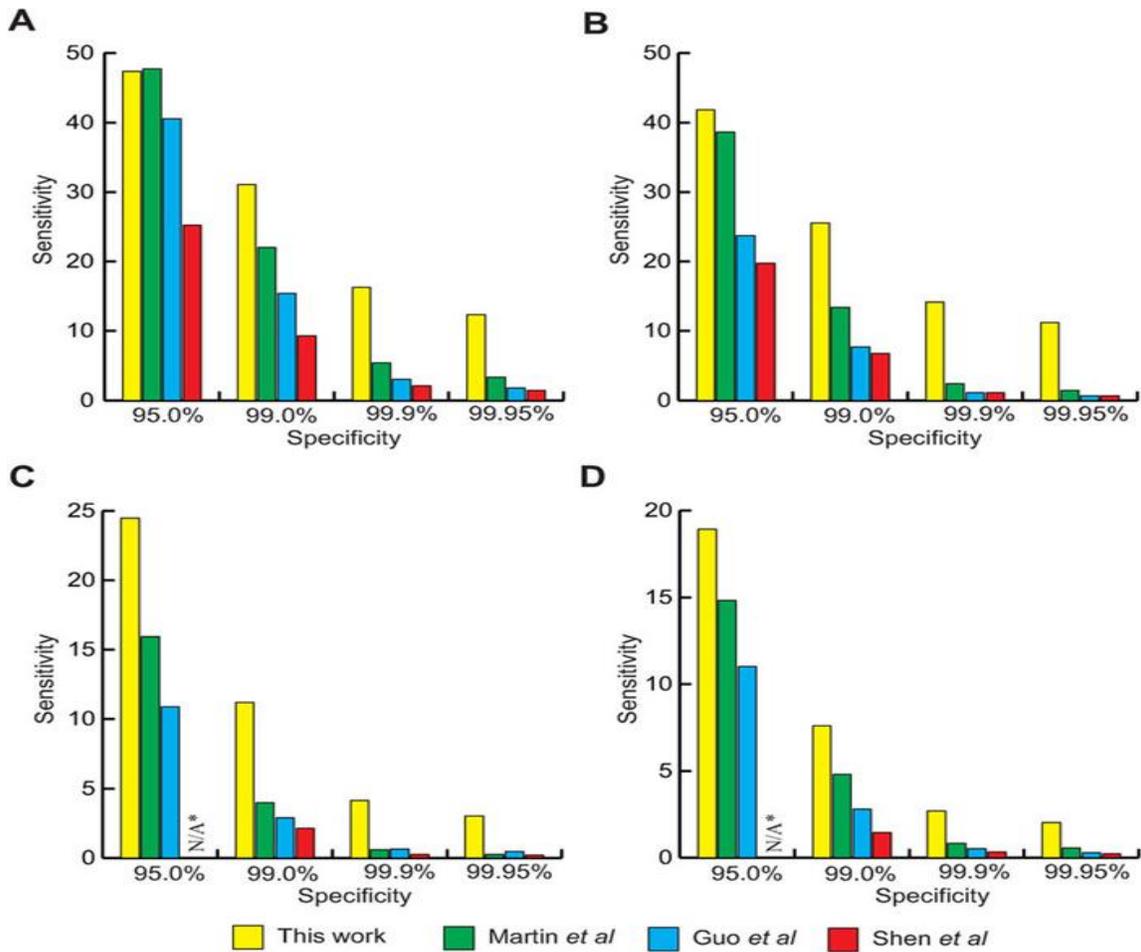


Figure 2.4: Accuracy vs. specificity for this work compared to previous work (Guo Y. et al. 2010, Shen J. et al. 2007 and Martin S. et al. 2005) as found in (Park Y. 2009). A. Using *H. sapiens* data to predict *H. sapiens* interactions. B. Using *S. cerevisiae* data to predict *S. cerevisiae* interactions. C. Using *H. sapiens* data to predict *S. cerevisiae* interactions. D. Using *S. cerevisiae* data to predict *H. sapiens* interactions. *The experiment by Park in (Park Y. 2009) did not have results for 95.0% specificity for Shen et al in (C) and (D): the results for this method jumped from 2-3% specificity to 97-98% with no intermediate values in both experiments.

for different organisms. Appendix 11 shows the difference between simply applying the unmodified PPI prediction method using yeast parameters to human proteins and the adapted human PPI prediction method using the PPI prediction adaptation algorithm described in Materials and Methods.

Global interactomes are expected to be highly imbalanced, where the vast number of protein pairs are not expected to form true interactions. A recent paper (Yu CY et al. 2010) compares their method to two other sequence-based prediction approaches (Guo Y et al. 2008 and Shen J et al. 2007) using an unbalanced dataset where there are significantly more negative pairs than positive pairs used for training. The paper evaluates balanced datasets (1:1 positive-to-negative ratio) up to a mildly unbalanced 1:15 positive-to-negative ratio comparing methods using a combined sensitivity and precision measure called the *F-measure*. Yu *et al* reported that the *F-measure* for all methods tested decreased as the positive-to-negative ratio increased. Using the same datasets published in (Yu CY et al. 2010) we tested our method using the most unbalanced dataset examined in the paper (1:15 positive-to-negative ratio). Note that the real positive-to-negative ratio in *H. sapiens* is expected to be as high as 1:500 or more, but Yu *et al* only examined up to 1:15 ratio due to technical limitations. Table 2.4 presents the results for the three methods published in (Yu CY et al. 2010) and also our own method. Using this unbalanced dataset our approach achieves a significantly higher *F-measure* of 62.9 ± 1.1 while the next-best method (Yu CY et al. 2010) yields a value of 43.6 ± 1.3 . It is noteworthy that our method also recorded better accuracy, precision, sensitivity and specificity than any other method tested.

Table 2.4: Comparison of this work with previous works (Guo Y et al. 2008, Shen J et al. 2007 and Yu CY et al. 2010) using the evaluation datasets with 1:15 positive-to-negative ratio compiled by Yu *et al.* The cutoff parameter was chosen to maximize the *F-measure* in order to compare results with those in Yu CY *et al* 2010.

	Acc. (%)	Fm (%)	Prec. (%)	Sens. (%)	Spec. (%)
Shen <i>et al</i>	92.5 ± 0.1	33.1 ± 1.4	37.5 ± 1.3	29.7 ± 1.5	96.7 ± 0.1
Guo <i>et al</i>	91.7 ± 0.2	36.6 ± 1.5	35.1 ± 1.5	38.3 ± 1.9	95.3 ± 0.2
Yu <i>et al</i>	93.7 ± 0.2	43.6 ± 1.3	49.5 ± 1.7	39.0 ± 1.3	97.3 ± 0.1
This work	95.7 ± 0.1	62.9 ± 1.1	73.7 ± 2.8	55.0 ± 1.6	98.6 ± 0.2

We have emphasized sensitivity and specificity as performance measures in this paper (as opposed to precision or *F-measure*) since they are independent of the ratio of positive to negative protein pairs in the test data. As illustrated in Table 2.2, precision depends directly on this ratio, and furthermore, the true ratio of interacting protein pairs is unknown for most organisms. Appendix 3 illustrates the precision of our method over a range of positive-to-negative ratios when operating at the threshold values given in Table 2.3. As expected, the precision is negatively impacted by a decreasing ratio of positive-to-negative test samples as the positive predictions become increasingly dominated by false positives. Appendix 4 compares our precision over human proteins compared to three competing methods examined in (Yu CY et al 2010) over a range of ratios from 1:1 to 1:1000. Note that this far exceeds the range of ratios examined in ref (Yu CY et al 2010). Our method clearly outperforms all other methods at all ratios. It should be noted that, unlike other algorithms based on machine learning, our method does not require training *per se*. The only parameter which must be tuned is the PAM sequence window similarity score threshold, which is determined using random sequences drawn from the target organism's proteome (i.e. not from the database of known PPIs; see Methods for details). All performance results are computed using a stringent leave-one-out protocol which ensures that the test data is independent from the database of known interactions on which the predictions are based.

2.4.4 Effects of homologous sequences

It may be expected that homologous sequences in our evaluation data sets may lead to overfitting of the method and a corresponding overestimate of prediction performance. We therefore investigated the impact of homologous sequences in our data. Following the approach taken by Park, for each organism all homologous sequences were removed from the

databases such that the remaining proteins share less than 40% sequence identity. The most dramatic effect was observed for human, where removing homologs reduced our sequence set from 22,513 to 14,867 proteins. This resulted in a corresponding decrease in our known interaction set from 41,678 to 19,588 pairs (47% of the original set). LOO analysis was repeated for each organism to compare the performance with and without homologous sequences. Results from these experiments are illustrated using ROC curves in Appendices 5-9. For human, at lower specificities we notice a drop in sensitivity of approximately 7-8%, however, at high specificities (required for all-to-analysis as discussed below) the reduction in sensitivity becomes less pronounced. *S. cerevisiae* was affected to a lesser degree (4% sensitivity decrease at 99.95% specificity) and in the other organisms (*C. elegans* and *S. pombe*) the results actually improved slightly at high specificities. Again, PPI prediction in *E. coli* seems to be an exception, suggesting differences between conservation of protein interaction codes in *E. coli*, a prokaryote, and eukaryotic organisms. Even with homologous sequences removed from our human interaction database we continue to perform as well as the methods presented in Table 2.4. This is despite the fact that the data sets used to train those methods (Park Y 2009) contain a large percentage of homologs (estimated to represent 55% of their overall positive data set when following the same homology analysis used here (Park Y 2009)).

2.4.5 Computational prediction of *S. pombe* and *C. elegans* global interactome

At this point, the run time for our approach still precludes analysis of the full human global protein interaction map (requiring ≈ 6.3 million CPU hours). We therefore targeted *S. pombe*, which has 5,024 proteins and has the second best ROC curve in our initial small scale LOO analysis. Large scale genome-wide analysis was conducted on all possible *S. pombe*

protein pairs. We detected a total of 9,009 possible interactions, 6,058 of which are novel and suitable for experimental validation (see Appendix 2.1). Since currently there are only 2,951 known interaction pairs in *S. pombe*, our predictions have potentially increased our knowledge of the *S. pombe* interactome by over three fold. To examine the quality of the predicted interaction pairs we classified them according to their molecular function, biological process, and sub-cellular location (Table 2.5). As indicated in Table 2.5(a) to (c), a significant portion of protein pairs predicted via interaction codes have similar functions (39.9%), occur in the same cellular component (61.8%) and participate in the same cellular process (34.5%). This is comparable to levels of agreement for previously reported protein pairs, which are 53.9%, 79.0%, 54.2%, respectively, and significantly higher than those for random pairs (2.4%, 32.6% and 7.3%, respectively). Third party interactions (where both partners interact with another common protein) were also investigated to further assess the quality of our predicted interactions. Again 36.2% of protein pairs predicted via interaction codes had a common third protein partner, compared to 76.5% for previously reported and only 0.032% for random pairs. It should be noted however that for these analyses (especially third party interaction), the previous experimentally detected interactions might have an unfair advantage since proteins known to interact are often prioritized for further analysis and characterization.

Following our *S. pombe* experiments, *C. elegans* was targeted as another model organism of interest. It has nearly five times more confirmed proteins than *S. pombe* (23,684 compared to 5,024 respectively). A genome-wide scan of *C. elegans* resulted in 37,572 possible interactions, 31,056 of which are novel predictions. We again classified those predictions according to their molecular function, biological process, and location inside the

Table 2.5: Percentages of *S. pombe*, *C. elegans* and *H. sapiens* pairs in which both partners share the same GO SLIM annotation as well as third party interactions. (A) Results for 100,000 random *S. pombe* pairs. (B) Results for 2,951 previously known *S. pombe* interactions from BioGRID. (C) Results for our 9,009 *S. pombe* predicted interactions. (D) Results for 100,000 random *C. elegans* pairs. (E) Results for 6,607 previously known *C. elegans* interactions from BioGRID. (F) Results for our 37,572 *C. elegans* predicted interactions. (G) Results for 100,000 random *H. sapiens* pairs. (H) Results for 41,678 previously known *H. sapiens* interactions from BioGRID. (I) Results for our 1,056 *H. sapiens* predicted interactions.

	Derived from GO annotation				Third Party Interaction
	Cellular Component (CC)	Molecular Function (MF)	Biological Process (BP)	CC & MF & BP	
(a) Random <i>S. Pombe</i> pairs	32.6%	2.4%	7.3%	1.0%	0.032%
(b) Previously detected <i>S. Pombe</i> interactions	79.0%	53.9%	54.2%	29.9%	76.5%
(c) Predicted <i>S. Pombe</i> Interactions	61.8%	39.9%	34.5%	18.9%	36.2%
(d) Random <i>C. elegans</i> pairs	1.1%	2.4%	4.8%	0.2%	0.08%
(e) Previously detected <i>C. elegans</i> interactions	9.7%	44.6%	48.5%	4.2%	28.8%
(f) Predicted <i>C. elegans</i> Interactions	4.7%	30.2%	30.2%	1.8%	6.8%
(g) Random <i>H. sapiens</i> pairs	31.7%	32.7%	28.2%	11.3%	0.3%
(h) Previously detected <i>H. sapiens</i> interactions	82.5%	90.1%	82.2%	67.7%	59.2%
(i) Predicted <i>H. sapiens</i> interactions	91.2%	92.9%	88.9%	83.2%	45.3%

cells as we did with our *S. pombe* predictions (Table 2.5(d) to (f)). For our predictions, the percentage of pairs that simultaneously have similar function, occur in the same cellular component, and also participate in the same cellular process is 1.8%, which is consistent with the percentage for previously reported protein pairs (4.2% for 6,607 pairs). In contrast, for randomly selected protein pairs, the percentage of pairs that have similar function, occur in the same cellular component and participate in the same cellular process is only 0.2% (for 100,000 tested random pairs). The results for similar function and cellular process are analogous to those of *S. pombe*. However, the values for cellular component do not appear to follow the same trend. This might be due to a lack of reliable data since even the experimentally determined PPIs do not show any enrichment. For proteins pairs predicted in *C. elegans*, 6.8% had a common third protein partner compared to 28.8% in previously reported pairs and only 0.08% in random pairs.

Note that our PPI prediction software is based solely on conserved interaction codes and does not use any information about a protein's molecular functions, biological processes or sub-cellular location. Therefore, such additional information about the biological activities of protein pairs can potentially also be utilized to determine an independent confidence level for a predicted interaction. Since true interactors are thought to be functionally related, participate in the same cellular process, occur in the same sub-cellular location, and interact with a same interacting partner (third partner), such information can be used to determine a confidence level for a predicted interacting pair. For example, a predicted protein pair "A" with both partners sharing the same 1. molecular function, 2. cellular process and 3. interacting partner, may have a higher confidence level than protein pair "B" with partners having the same 1. molecular function and 2. sub-cellular location, which may have a higher

confidence level than protein pair “C” with partners only being co-localized in the same sub-cellular location. It should be noted however, that this approach may discriminate against uncharacterized proteins for which limited information is available.

2.4.6 Prediction of novel protein complexes

PPI data can also be used to determine protein complexes. From our predicted novel PPIs we identified two novel complexes for *S. pombe* membrane proteins. Membrane proteins often provide a challenge for classical PPI detection methods. The first is a five member complex shown in Figure 2.5A with SPBC1289.13c, a putative galactosyltransferase, as a core protein that interacts with four other proteins, some of which also interact with each other. Two of these proteins, SPAC22E12.06c and SPCC736.04c, are known to have alpha-1,2-galactosyltransferase activities involved in N-linked, and both O-linked and N-linked oligosaccharide modification of proteins, respectively (Yoko-o T et al. 1998), and have previously been reported to interact with each other (previously reported PPIs are shown in green). All five proteins are membrane proteins that are associated with the Golgi apparatus. Therefore it is likely that this complex has a role in galactosylation of glycoproteins. The second complex shown in Figure 2.5B consists of 8 members, with 5 proteins forming the core, and 3 additional proteins that interact with the core proteins but not with each other. They are all thought to be membrane proteins. Five of these proteins (SPBC1703.10, SPAC4C5.02c, SPAC6F6.15, SPAC9E9.07C and SPAC18G6.03) have been linked to protein transport and vesicular trafficking. This suggests a role for the complex in this process.

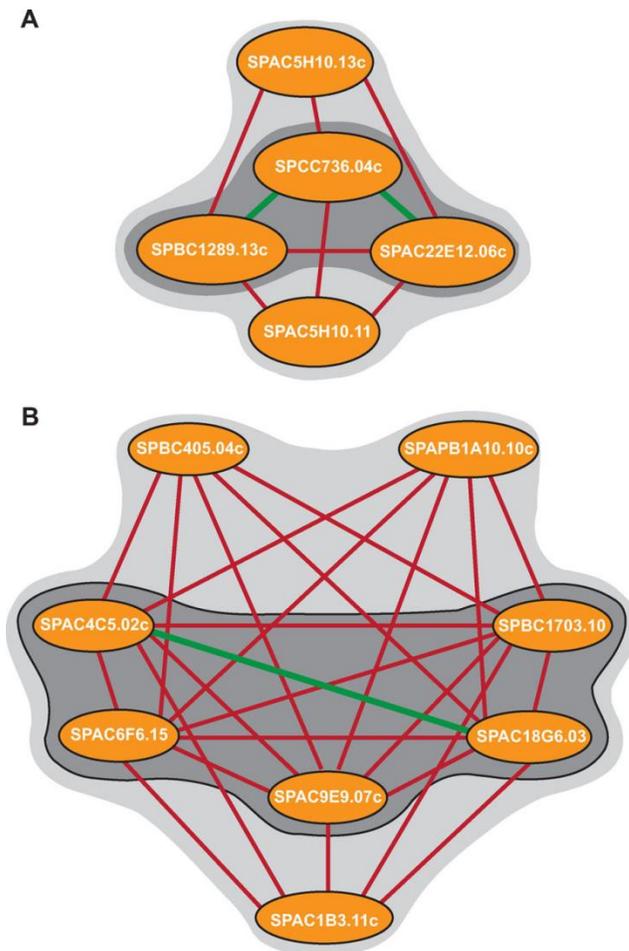


Figure 2.5: (A) A novel five member complex for *S. pombe* identified using interaction codes. Previously reported PPIs are shown in green; novel PPIs are shown in red. Previous annotations of key proteins suggest that the complex may play a role in galactosylation of glycoproteins. The dark grey shade represents core proteins. (B) A novel eight member complex for *S. pombe* formed by 5 fully-connected core proteins (shaded) and three peripheral proteins. Predicted PPIs combined with previous annotation of some proteins suggest putative function in protein transport and vesicular trafficking. The dark grey shade represents core proteins.

2.4.7 Investigation of human dsDNA break repair PPI network

To evaluate the effectiveness of interaction codes to predict PPIs in human we investigated the interactions for 29 proteins, with established roles in the efficiency of double stranded (ds) DNA break repair, against all human proteins. This represents an analysis of more than 650,000 possible protein pairs (29 DNA break repair proteins against all 22,500 reviewed *H. sapiens* proteins from Uniprot (Jain E et al. 2009). dsDNA breaks represent a severe case of DNA damage which can lead to cancer development if left unrepaired. It is estimated that about 1% of proteins in the genome function in the DSB repair process. We detected a total of 620 interactions (at 99.95% specificity), 349 of which were previously reported, and 271 of which represented novel interactions (see Appendix 2.2). This represents an expansion of the known dsDNA break repair interactome by an additional 75%.

Following the PPI prediction verification analysis used above, we again observed (Table 2.5 (g) to (i)) that a significant portion of our predicted partners have the same molecular function (92.9%), are involved in the same cellular processes (91.2%) and are co-localized (88.9%), highlighting the quality of our predictions. Interestingly our predicted interacting partners had more in common with each other than previously reported partners (90.1%, 82.5% and 82.2%, respectively). This is consistent with the ROC curves which show better prediction abilities in human than in *S. pombe*. Predicted PPIs are further verified when third party interactions are examined: only 0.3% of random protein pairs had third party interactions, while this number rose to 45.3% in predicted PPI pairs – much closer to the 59.2% observed among previously reported PPIs.

An interaction between two proteins often infers a functional relationship between the two. We next examined how our predicted PPIs can be used to infer novel protein function.

We hypothesized that 8 proteins Q9NNZ6 (PRM3), Q92769 (HDAC2), Q9NRF9 (POLE3), P31946 (YWHAB1), Q17RH7 (TPRXL), Q15072 (ZNF146), Q9NVP2 (ASF1B), and Q5EBM4 (ZNF542) may be involved in double stranded DNA (dsDNA) break repair due to their novel interactions with proteins known to be involved in this process (Fig. 2.6A). To test this hypothesis we experimentally investigated the activity of these proteins. This was done by subjecting the yeast gene deletion mutants for corresponding human homologs to a plasmid repair assay (Fig. 2.6B). It was observed that deletion of 4 (*YPL009C*, *HDA1*, *MNN4*, and *BMH2*) of the 8 genes (or 50% of the tested genes) resulted in a statistically significant (P -value < 0.05) decrease in the ability of the mutant cells to repair dsDNA breaks. This suggests an involvement for their corresponding human genes *PRM3*, *HDAC2*, *POLE3*, and *YWHAB1*, respectively in the efficiency of DNA damage repair. Due to technical limitations associated with our plasmid repair approach, it is likely that more of the tested candidate genes may, in fact, be involved in dsDNA break repair. Their deletion may have subtle (or no) phenotypic consequences that cannot be detected by our current experimental method.

Some of the novel proteins in Figure 2.6A, for example, Q9NNZ6 and Q5EBM4 form numerous interactions with the known dsDNA break repair proteins, whereas others, for example, Q17RH7 and P31946 form a single interaction each. It is generally thought that the degree of connectivity of a protein within a PPI network may represent the significance of that protein within the system. Consequently, the “hub” proteins are thought to be essential for the integrity of a PPI network. In agreement with this, 3 of the 4 identified novel proteins are highly connected and only one novel protein, P31946, forms a single interaction.

Interestingly, deletion of its homolog in yeast (*bmh2Δ* stain) had the lowest effect on the efficiency of dsDNA break repair among the 4 novel positives.

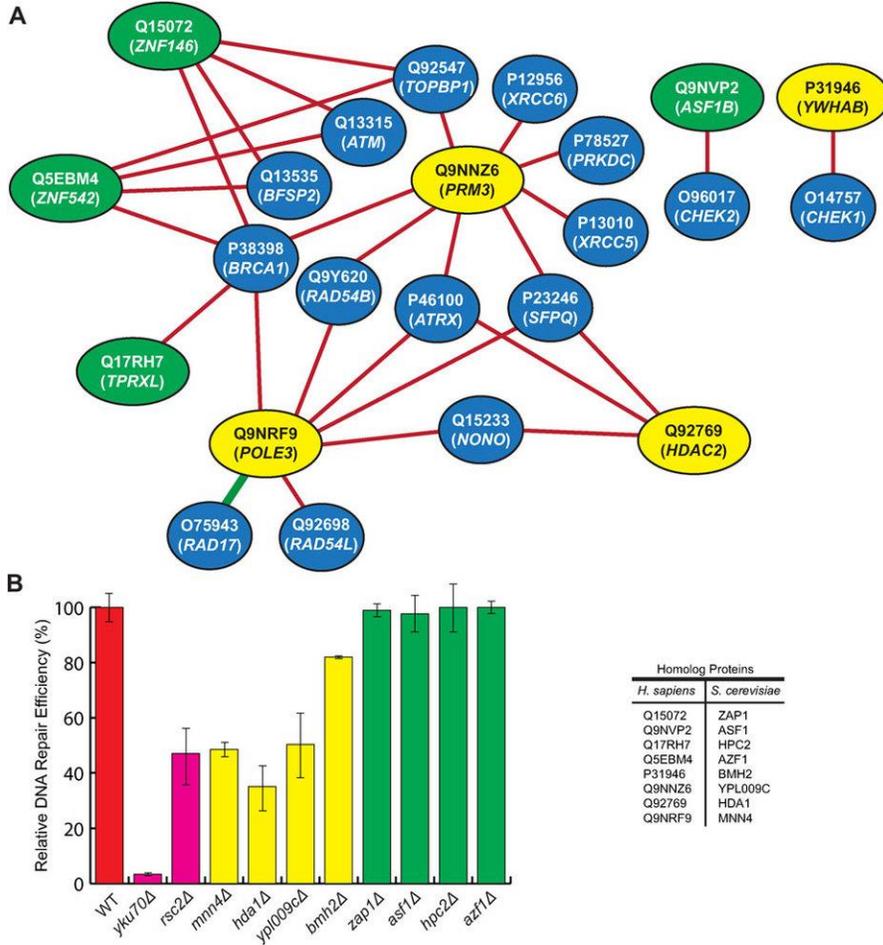


Figure 2.6: Inferring protein function from predicted PPIs. Biological process may be inferred based on the annotations of predicted interaction partners. (A) Predicted PPIs for 8 novel human proteins (yellow and green nodes) against known dsDNA break repair proteins (blue nodes). Red edges represent novel predicted interactions and the green edge represents a previously reported interaction. (B) Yeast deletion mutant strains for genes homolog to human novel genes above are subjected to plasmid repair assay. The number of colonies formed after strain transformation with linearized plasmid is normalized to that of intact plasmid, and related to the wild type (red bar) strain (set at 100%). *YKU70* and *RSC2* are known players in non-homologous repair of double stranded DNA breaks and are used as positive controls (pink bars). Yellow bars represent strains with statistically significant (P -value < 0.05) reduction efficiency in plasmid repair. The strains that did not show reduction in plasmid repair efficiency are represented by

2.5 Discussion

Through computational acceleration of our PPI prediction algorithm based on conserved interaction codes, we have performed the first genome-wide PPI analysis for *S. pombe* and *C. elegans*. The new computational abilities enabled us to demonstrate the conservation of interaction codes among multiple species and that prediction of PPIs in an uncharacterized proteome can be performed based on known PPIs from other species. Furthermore, the human dsDNA interactome was considerably expanded through our method. This analysis led to the assignment of novel protein functions which was confirmed experimentally for four proteins.

The current algorithm works on the basis of available PPI data. In addition to being sparse, the available PPI data contain numerous false positives and hence are considered “noisy” data. One may expect that the availability of additional high confidence PPI data may help increase the performance of the current approach. Advances in protein 3D structures can be related to the growing number of structures for interacting complexes to predict novel binding partners. Growing databases of computationally predicted PPIs can reveal new information about novel properties of interacting partners. In future, some of these properties may be applied to the current algorithm to reduce false positives and hence increase specificity. GO terms can also be applied to eliminate false positives. These improvements however, may come in expense of novel predictions since they can discriminate against uncharacterized proteins.

The origin of the interaction codes used here for PPI prediction is not clear. They seem to be present in both eukaryotes as well as in *E. coli*, a prokaryote. However, the fact that

eukaryotic codes cannot accurately predict *E. coli* PPIs may suggest a difference between pro and eukaryotic codes. A possible explanation for the origin of these short polypeptide regions is that they may have evolved from longer interaction mediating domains in order to maximize sequence usage. These codes represent alternative polypeptide signals that merit further attention.

The ability to investigate PPIs at a proteome level sets path to better study the topology of different PPI networks. In this context “centrality” of a network can be better elucidated by examining hub and betweenness centralities (Jeong H et al 2001 and Yu H et al. 2007). These values can indicate proteins thought to be more relevant to the integrity of a network. They can provide important clues for disease progression. They can also prove to be important drug targets in different pathogens.

In addition to the applicability of the new algorithm to study arbitrary species, the new algorithm is also considerably faster. The increased speed associated with the current algorithm however, is not sufficient for studying human genome wide PPI analysis. Using Massive Parallel (MP) computing approaches, we are currently in the process of generating a new method amenable to studying human PPI network.

Chapter 3: *TPK1* influences the efficiency of non-homologous end joining

3.1 Abstract

Non-homologous end joining (NHEJ) is a highly conserved mechanism of DNA double-stranded break (DSB) repair. Here we show that the deletion of yeast *TPK1* that codes for the protein kinase A catalytic subunit reduces efficiency of NHEJ repair of breaks with overhangs and blunt ends in plasmid based repair assays. Our double deletion mutant analyses suggest that *TPK1* and *NEJ1* work in the same pathway and in parallel to *YKU80*. Additionally, *tpk1Δ* mutants showed defects in the repair of chromosomal breaks induced by HO-site specific endonuclease. This deficiency was compensated by *NEJ1* overexpression. Chromatin immunoprecipitation (ChIP) analysis further indicated that the recruitment of Nej1 to the site of a DNA break is influenced by Tpk1. The effect of Tpk1 on NHEJ efficiency seems to be conserved between yeast and mammalian systems, since mutants of PRKACB, the human homolog for Tpk1, reduced the efficiency of NHEJ in an *in vitro* DNA repair assay. Altogether, here we report a novel involvement for Tpk1 in NHEJ.

3.2 Introduction

DNA double stranded breaks (DSBs) are the most severe forms of DNA damage. DSBs are induced either endogenously by replication errors and metabolic oxidative products or as a result of exogenous reagents such as environmental stress, radiation or toxic chemicals (Schipler A and Iliakis G 2013). Once the phosphodiester bond of both strands is damaged, the DNA will be divided into two fragments. This fragmentation can cause genomic instability and chromosomal rearrangement. Mutations arising from mis-repaired or unrepaired DSBs can lead to the onset of complex diseases such as cancer,

neurodegeneration and immunodeficiency (Watering E and Chen DJ 2008). DSBs are primarily repaired by two pathways that work independently: homologous recombination (HR) and non-homologous end joining (NHEJ) (Symington LS and Gautier J 2011). HR uses homologous templates usually obtained from sister chromatids or homologous chromosomes to connect the broken sites of DSBs, producing an error-free repair. This pathway is preferred in cases of stalled replication forks (Saleh-Gohari N et al. 2005). DSBs can also be ligated directly through NHEJ pathway in an error-prone manner (Churuvella KK et al. 2013). NHEJ is the main repair pathway in human cells and is highly conserved from simple eukaryotes such as the yeast, *Saccharomyces cerevisiae*, to a more complex mammalian system (Deriano L and Roth DB 2013).

The process of NHEJ involves three main protein complexes for repair initiation, ligation preparation and, finally, for the rejoining of the broken ends. The first protein complex to initiate NHEJ is Yku70/Yku80, which attaches to the broken ends of DNA and recruits other repair factors to the site of damage (Leiber MR 2010). Recognition of the damage and recruitment of the YKU complex to the site of damage involves a cascade of DNA damage checkpoints in which Tel1, Mec1 and Rad53 play central roles (Ataian Y and Krebs JE 2006). Once YKU is bound to the DNA ends, it recruits the MRX complex (Mre11, Rad50 and Xrs2) to the site of damage (Wu D et al. 2008). In preparation for ligation, the MRX complex forms a bridge between the broken ends and brings them in close proximity (Williams RS et al. 2008). The final ligation of the DSB requires Lif1/Lig4 complex. Lif1 is brought to the break site by YKU and subsequently recruits Lig4, a protein with ligase activity (Palmbos PL et al 2008). Another factor suggested as part of Lif1/Lig4 complex is Nej1 due to its interaction with Lif1 (Deshpande RA and Wilson TE 2007). The C-terminal

region of Nej1 plays a critical role in its function in NHEJ. This region is the site of interaction with Lif1, and has DNA binding affinity. Mutations within the C-terminus of Nej1 have been shown to reduce NHEJ efficiency to the same level as *NEJ1* deletion (Mahaney BL et al. 2014). Even though Nej1 does not interact directly with Dnl4, its recruitment to the site of damage is dependent on the presence of Dnl4 (Wu D et al. 2008). Additionally, Nej1 gets phosphorylated by Dun1 at serine 297 and 298 in response to DNA damage, though the influence of this phosphorylation in the role of Nej1 in NHEJ is not clear (Ahnesorg P and Jackson SP 2007 and Chovanec M and Wilson TE 2007). Nej1 is also shown to affect NHEJ via YKU independent pathways such as micro-homology mediated end joining (MMEJ) (McVey M and Lee SE 2008). MMEJ, which requires Nej1, plays a significant role in the repair of DNA breaks with blunt ends (Lee K and Lee SE 2008). The YKU heterodimer is insignificant in the repair of DNA breaks with non-cohesive ends (Boulton SJ and Jackson SP 1996 and Yu L and Volkert MR 2013). Even though the interactions within the core components of NHEJ are well understood, several other proteins have been shown to play a role in NHEJ such as Rtt109 (Jessulat M et al. 2008), Tdp1 (Bahmed K et al. 2010) and Pph3/Psy2 (Omid K et al. 2014), suggesting existence of other proteins in the NHEJ pathway.

Kinases often play a central role in signal transduction and protein networking within a pathway and between different pathways in the cell. Protein kinase A (PKA) is one of the primary kinases functioning in many cellular processes such as metabolism, stress response, nutrient starvation and the cell cycle (Galello F et al. 2010). Due to its diverse functions, PKA has been targeted in studies concerned with different areas of research including anti-aging (Enns et al. 2010) and drug discovery (Åberg E et al. 2012). PKA has been reported to

function in mitosis by regulating anaphase progression (Anghileri P et al. 1999). Another study has connected the role of PKA in mitosis to DNA damage response (Searle JS et al. 2004), proposing two different models. It was suggested that PKA is either activated in response to cell cycle arrest and abundance of cell cycle regulatory proteins such as Cdc20, Pds1 and Clb2, or it is directly activated by Mec1, a DNA damage checkpoint protein (Searle JS and Sanchez Y 2004). In yeast cells, PKA has three catalytic protein subunits, Tpk1, Tpk2 and Tpk3 (Searle JS et al. 2004). Even though the role of PKA in the DNA damage response remains unclear, it is suggested that its catalytic subunits Tpk1, Tpk2 and Tpk3 may play important roles in this pathway (Searle SJ et al. 2011).

Here we report a role for Tpk1 in DSB repair through NHEJ. Our data shows that the deletion of *TPK1* significantly reduces the efficiency of NHEJ repair in both plasmid and chromosomal break repair assays. We demonstrate that Tpk1 alters the efficiency of NHEJ repair through its association with Nej1.

3.3 Material and Methods

3.3.1 Yeast strains and plasmids

Yeast mutant strains were obtained from *S. cerevisiae* deletion library in BY4741 background (MATa orf Δ ::KanMX4 *his3* Δ *leu2* Δ *met15* Δ *ura3* Δ) (Winzler EA et al. 1999). *TPK1* was deleted in BY4741 and Y7092 (MATa *can1* Δ ::STE2pr-HIS3 *lyp11* Δ *ura31* Δ *leu21* Δ *his31* Δ *met151* Δ) background by PCR transformation method as in (Tong AH and Boone C 2007), which replaces the target gene with a Nat^R cassette. Gene deletion strains in JKM139 (MATa *hmr* Δ ::ADE1 *hml* Δ ::ADE1 *ade1-100* *leu2-3,112* *lys5* *trp1*::*hisG* *ura3-52* *ade3*::GAL-HO) background were constructed as in (Moore JK and Haber JE 1996). Single

and/or double mutant strains in JKM139 were constructed through transformations using PCR products containing Nat^R and/or Hyg^R to replace target gene(s). Plasmid p416 which has a URA3 marker and Amp^r gene (Jessulat M et al. 2008) was used for plasmid repair assay. Ycplac111 plasmid with Leu2 marker and Amp^r was used to test repair efficiency of plasmid carrying a blunt end break as in (Bahmed K et al. 2010). Suppression analysis was carried out using different gene expression plasmids in pGEH background under the transcriptional control of Gal1 promoter with a 2 micron origin of replication as explained (Sopko R et al. 2006).

3.3.2 Cell culture, transfections, and PRKACB knock out (KO) construction

U2OS cells were cultured and maintained with Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum. *PRKACB* KO construct was generated in CRISPR-Cas9 system, using lentiCRISPR ver 2 (obtained from Addgene) following the established procedure (Shalem O et al. 2014). The single guide RNA (sgRNA) oligonucleotides (AATGTGCCTGTGCCACCAGA) targeting to *PRKACB* was selected from the large-scale human CRISPR-Cas9 KO library generated by Zhang's group using the CRISPR design tool (Shalem O et al. 2014). The chemically synthesized sgRNA oligonucleotides of *PRKACB* were cloned into the LentiCRISPR backbone containing ampicillin resistance marker. Approximately 2.5 µg of the sequence verified plasmid DNA (i.e., LentiCRISPR containing *PRKACB* sgRNA) was transfected into the cultured U2OS cells in 6 well plates using standard Lipofectamine® LTX plus transfection procedure. Confirmation of *PRKACB* knockout was performed via sequencing. Genomic DNA was isolated from empty vector- or sgRNA-directed *PRKACB* knockout-transfected U2OS cells according to manufacturer's instructions (Qiagen). Phusion high-fidelity PCR (Thermo

Scientific) was used to amplify a genomic locus, flanking exon 7 of *PRKACB* gene (binding target of sgRNA) with 150 ng gDNA input along with following primers: forward primer 1: 5'-GAAATTTACGATGAGGGATCTTCA-3', and reverse primer 1: 5'-GGGAGGAAGGAGGGAAAG-3'. PCR product was gel-purified using E.Z.N.A gel extraction kit (Omega Bio-Tek), and the resulting purified PCR mixture was Sanger-sequenced with the aforementioned forward primer 1 and a nested forward primer 2 (5'-TGTCATGGAAG AGTTACA AAA-3'). Sequencing results were then aligned and visualized using Jalview 2.8.2.

3.3.3 Protein-protein interaction prediction

Prediction of protein-protein interactions (PPIs) were based on short co-occurring conserved polypeptide regions as described in (Pitre S et al. 2008) with the following modifications. An updated list of high confidence human PPI interactions was gathered from published data (BioGRID: www.thebiogrid.org and DIP: www.dip.doe-mbi.ucla.edu). Sensitivity (the percentage of interactions that can be identified from the starting data set) was increased by a fold to an approximately 45% by reducing specificity (a measure of false positives) to 90%. Regions that mediate PPI between proteins were predicted as described in (Amos-Binks A et al. 2011).

3.3.4 Plasmid repair assay

Circular plasmids p416 and Ycplac111 were digested at their unique XbaI and SmaI restriction sites within regions with no homology to yeast chromosomes, respectively. An equal amount of circular and linearized plasmids were transformed into single mutant, double mutants and WT strains using lithium acetate transformation protocol. Cells were grown on

minimal media (2% agar, 2% glucose, 6.74g yeast nitrogen base/L, 2g amino acid supplement without uracil or leucine /L) at 30° C for 2-3 days. The ratio of colony numbers of linearized to circular plasmid transformations was normalized to that of WT. Each trial was repeated at least five times.

3.3.5 Drug sensitivity tests

Yeast strains were grown in YPD at 30°C to saturation then serially diluted by factors of 10. 15ul of each dilution was spotted on YPD media containing 3 ul/ml bleomycin or 70 mM hydroxyurea (HU) and drug free media. Cells were grown at 30°C for 1-2 days and growth was visually examined.

3.3.6 Synthetic genetic array analysis

To study genetic interactions for *TPKI*, we employed synthetic genetic array analysis described in (Tong AH and Boone C 2006). Briefly, *TPKI* deletion mutant strain or its expression plasmid in alpha mating type (Y7092) was crossed with a DNA damage array (DDA) containing 384 single mutant strains associated with DNA repair pathways, cell cycle progression, checkpoints, chromatin modification or those genes whose products are found in association with chromatins, and an array of 384 random mutants of genes involved in various pathways as a control. Conditional dependent interaction analyses were performed in the presence of DNA damaging drugs bleomycin or HU. Fitness was measured by colony size. Quantification and analysis of colony growth was performed as in (Memarian N et al. 2007). Each experiment was repeated three times and the average values are presented. Relative growth difference of 20% or more in two repeats were considered positives.

3.3.7 Chromosomal break assay

JKM139 strain carries a GAL promoter linked to its HO-site specific endonuclease which is activated in the presence of galactose and produces DSBs at its MATa locus. Homologous regions, HML and HMR, to MATa are deleted in this strain. JKM139 WT, mutants and those carrying overexpression plasmids were grown in YPD or selective media for 2 days before being serially diluted to concentrations of 10^{-1} to 10^{-4} . Each dilution was spotted on YP-Galactose to induce double stranded breaks and YPD as a control. Growth of the mutant colonies was compared to that of wild type JKM139 after 2-3 days. Additionally each dilution was plated on media containing galactose or glucose and grown for 2-3 days. To quantify efficiency of chromosomal repair, the ratio of colonies formed on galactose medium to glucose medium for each mutant was normalized to that of WT. To calculate the p-values two-tailed distribution t-test was used.

3.3.8 Protein purification and *in-vitro* fluorescence kinase assay

Protein purification was done as described in (Sopko R et al. 2006). In brief, yeast cells were grown to $OD_{600} \sim 0.7$ at $30^{\circ}C$. GST-tagged proteins were affinity purified from yeast cell lysates, disrupted by glass beads in lysis buffer containing zymolyase and protease inhibitor cocktail, using glutathione agarose beads. The beads were washed five times with the wash solution containing 300 mM NaCl. Captured proteins were eluted from the beads by addition of the elution buffer containing 5 mM reduced glutathione. The purity of the purified proteins was estimated using SDS–polyacrylamide gel electrophoresis followed by Coomassie blue staining. An *in vitro* fluorescence kinase assay, ADP-Quest provided by DiscoverX, was used to measure kinase activity of Tpk1 (Wu ZL 2011). In brief, 3.5 ng or 7 ng of Tpk1 was incubated with 10 ng of Nej1, buffer A (provided by ADP-Quest kit) and ATP (from Bioshop). Fluorescence signal was detected as a result of ADP production and

measured at 530 nm. Experiments were performed at 30°C. Each data point represents the average of at least four independent experiments.

3.3.9 Chromatin immunoprecipitation analysis

Chromatin immunoprecipitation (ChIP) analysis was performed as described in (Faucher D and Wellinger RJ 2010) with the following modifications. A 40ml culture of JKM139 cells (Nej1:5xFLAG, Tpk1:5xFLAG, or Δ tpk1 Nej1:5xFLAG) were grown in YPD media overnight before being harvested by centrifugation, washed 3x in sterile water, and resuspended in YP-GAL media. 5ml samples were taken at the appropriate time points and cross-linking was done adding 1% formaldehyde for 30 minutes before crosslinking was terminated with 150mM glycine. To immunoprecipitate tagged protein and DNA, lysate was incubated with 3ul anti-DYKDDDDK antibody (Abcam) and 50ul protein G beads. RT-PCR was conducted using iQ SYBR Green supermix (Biorad) according to manufacturer's instructions, using primers 5'-TCACAGGATAGCGTCTGGAAGTCA-3' and 3'-CATTTTGTCCGCGTGCCATTCT-5' as in (Faucher D and Wellinger RJ 2010) .

3.3.10 NHEJ reporter assay in mammalian cells.

The NHEJ reporter construct (pEGFP-Pem1-Ad2) provided by V. Gorbunova (University of Rochester) was digested with I-SceI homing enzyme. The purified pre-digested product was transfected either with the empty vector control LentiCRISPR or PRKACB CRISPR KO constructs into the U2OS cells using Lipofectamine® LTX plus transfection. After 72 h of post-transfection, cells were harvested and fixed for FACS analysis. The percentage of GFP-positive cells in the KO and empty vector control constructs was calculated to measure the NHEJ repair efficiency.

3.4 Results and Discussion

3.4.1 Identification of TPK1 as a candidate involved in double stranded DNA break repair

Protein-protein interactions (PPIs) have been used extensively to better define the functions of various proteins (Dittrich MT et al. 2008). Recently, it was reported that a subset of human PPIs can be predicted on the basis of short co-occurring polypeptide sequences (Pitre S et al. 2012). Using this computational approach, we investigated the potential interactions that DNA repair proteins may form with other human proteins at a proteome level. In this context, PRKACB was identified as a protein of potential interest with a predicted ability to interact with three well-defined proteins involved in the process of DSB repair (Fig 3.1A). Its interaction list included KU70, a key protein in NHEJ repair initiation, DNA PKcs, which regulates the end processing of DNA in NHEJ, and ATR, a protein kinase involved in DNA damage response (Maréchal A and Zou L 2013) (Fig. 3.1A). SFPQ and NONO were also found in the interaction list of PRKACB; these proteins are involved in the DSB response and have been shown to interact with NHEJ proteins (Salton M et al. 2010). PRKACB has 49% sequence identity to the *S. cerevisiae* Tpk1, which constitutes the core component of the catalytic subunit of the PKA complex. On the basis of the PPIs that PRKACB formed in human cells we hypothesized that its yeast homolog, Tpk1, is a possible player in the NHEJ pathway, and subjected it to further investigation.

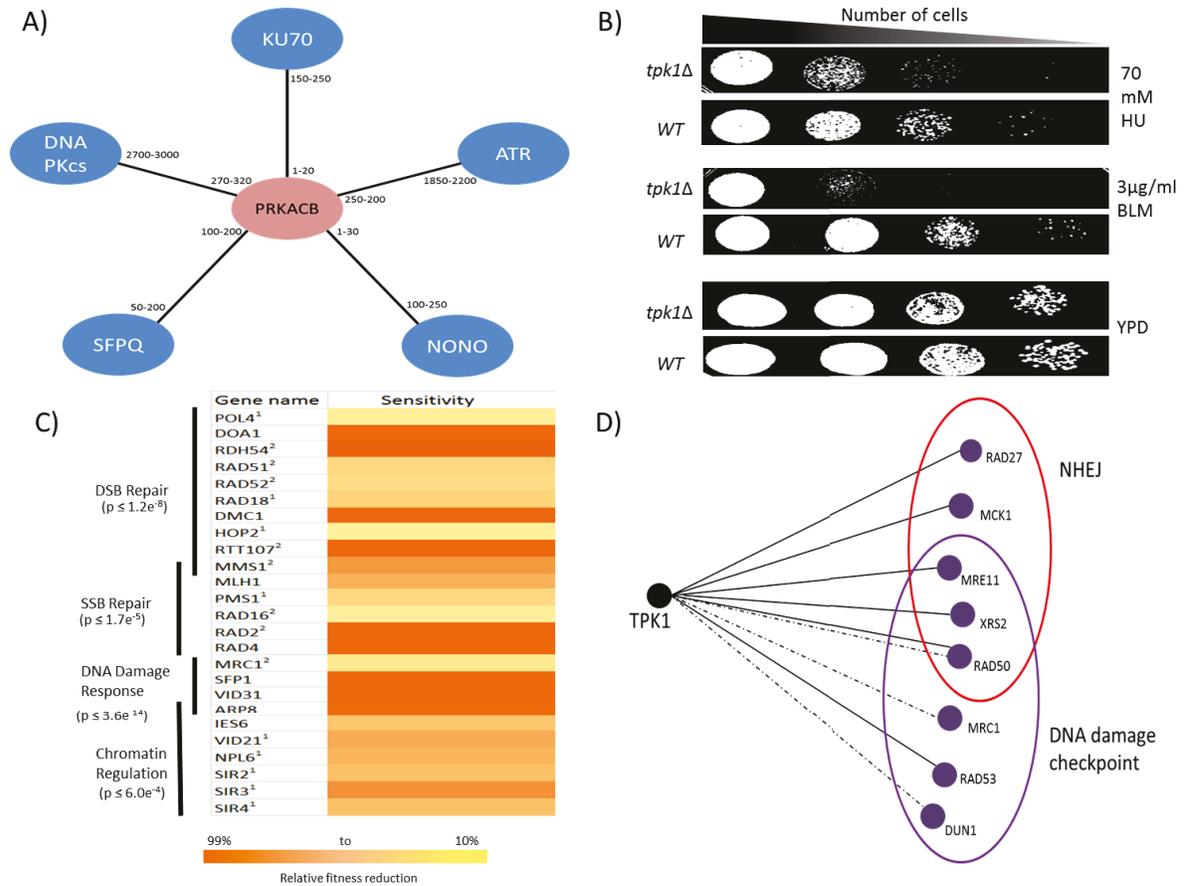


Figure 3.1: A) PRKACB is predicted to interact with several key NHEJ proteins including KU70. Numbers refer to amino acid residues and represent the predicted sites of interaction for interacting proteins. B) Deletion of *TPK1* makes the cells sensitive to bleomycin (3 $\mu\text{g}/\text{ml}$) and HU (70 mM). Cell survival is represented as spot tests. C) SGA and suppression analysis of *TPK1*. A) *TPK1* formed negative genetic interactions with a number of genes involved in DNA damage response and both SSB and DSB repair pathways. ¹ and ² refer to growth in presence of sub-inhibitory concentrations of bleomycin (1 $\mu\text{g}/\text{ml}$) and HU (40 mM), respectively. Relative growths are color coded as indicated B) Plasmid-borne *TPK1* expression suppresses deletion of genes involved in NHEJ and DNA damage checkpoint. Dashed and straight lines represent compensation of *TPK1* expression for sensitivity of gene deletion strains to bleomycin and HU, respectively.

3.4.2 *TPK1* deletion causes sensitivity to DNA damaging drugs

DNA damaging drugs induce different forms of damage at random locations on the DNA. Sensitivity to DNA damaging drugs is observed upon deletion of genes involved in different repair mechanisms (Billell GW et al. 2001). Hydroxyurea (HU) and bleomycin are two chemical agents that can produce DSB via distinct mechanisms. HU induces breaks by stalling replication forks through the depletion of dNTPs, causing replication stress (Koc A et al. 2004). Bleomycin is thought to generate free radicals after binding to DNA resulting in DSBs (Moore CW 1988). Deletion of *TPK1* made the cells extremely sensitive to low doses (3µg/ml) of bleomycin, while moderate sensitivity was observed in case of HU at an adequate (70 mmol) concentration (Fig. 3.1B).

3.4.3 Genetic interaction analysis of *TPK1*

Genetic interaction (GI) analysis often reveals functional redundancy for genes and pathways as well as the higher order of communication between different processes (Bandyopadhyay S et al. 2010). Consequently, it is commonly used to study novel gene functions (Costanzo M et al 2011). When the deletion of two genes causes a more severe phenotype than expected from individual gene deletions alone, it is said that the two genes form a negative GI also called synthetic sickness. Such aggravating interactions are generally explained by the activity of the two genes in parallel pathways. To study the activity of *TPK1* in DNA damage repair we investigated the negative interactions that its deletion made with two arrays of 384 deletion strains. One of these arrays, which we termed DNA damage array (DDA), contains deletion strains for genes that play a role in various DNA damage repair pathways, cell cycle progression, checkpoints, chromatin modification or those genes whose

products are found in association with chromatin. The second array, used as a control, contains deletion strains for random genes. Under standard laboratory growth condition, very few interactions were observed for *TPK1* (Fig. 3.1C). However when DNA damage was induced we observed that *TPK1* negatively interacted with a number of genes involved in DSB repair including *RAD51* and *RAD52* which are part of HR pathway and *RAD18* and *DMC1* which are involved in post-replication repair (Fig. 3.1C). It also interacted with genes associated with nucleotide excision repair (*RAD4*) and mismatch repair (*MLH1*). However it did not form any negative GIs with NHEJ genes. This pattern of negative interactions is very similar to those for well-defined NHEJ genes such as *YKU70*, *YKU80*, and *DNL4* (Koh JL et al. 2010). This can be explained by the fact that negative GIs are often formed between genes involved in parallel but not the same pathways (Boone C et al. 2007). Conditional dependency of the interactions (Bandyopadhyay S et al. 2010) that *TPK1* forms with DNA repair/response genes suggests that the activity of *TPK1* in DNA repair is dependent on the presence of DNA damage.

GIs can also be studied through suppression analysis in which the overexpression of a target gene compensates for a phenotype caused by the deletion of a second gene. This type of phenotypic rescue is termed dosage suppression (Magtanong L et al. 2011) and often reveals genes with similar activities. To this end, *TPK1* expression plasmid was introduced into the DNA damage and random arrays described above. It was observed that a sick phenotype to DNA damage caused by bleomycin and HU was suppressed by a plasmid-based expression of *TPK1* for a number of interesting kinases involved in DNA damage checkpoint including those encoded by *MRC1*, *RAD53* and *DUN1* (Fig. 3.1D). *TPK1* expression also rescued the phenotype of deletion of each member of MRX complex which plays a key role

in the DNA damage response, HR and NHEJ. These observations further connect *TPK1* to DNA damage repair. Other members of NHEJ genes that showed dosage suppression in our screen are *RAD27* and *MCK1*.

3.4.4 Deletion mutant for *TPK1* reduces the efficiency of plasmid repair through NHEJ

To study a possible involvement for Tpk1 in DSB repair, we subjected its gene deletion mutant strain to plasmid repair assays designed to measure the efficiency of NHEJ (Jessulat M et al. 2008 and Yu L and Volkert MR 2013). The plasmid based repair assay is a classical assay for the identification of novel genes involved in NHEJ (Boulton SJ and Jackson SP 1996). This assay can readily determine the repair efficiency of different types of DSB, thus providing an advantage in studying DNA breaks with cohesive ends as well as with blunt ends (Yu L and Volkert MR 2013). In these assays, plasmids are digested at a unique site with no homology to the yeast genome. Linearized plasmids with overhangs or blunt ends as well as intact plasmids are transformed into mutant and control strains in parallel. The ratio of colonies obtained from linearized and intact plasmids is normalized to that in a control strain and serves as a measure of the efficiency of plasmid dsDNA break repair using NHEJ. For *tpk1Δ* the repair efficiency of linearized plasmids with four nucleotide overhangs was measured at 12% (Fig. 3.2). This is comparable to that for *yku80Δ* (5%) used as a positive control. Of interest however, is that in case of blunt end digests, the repair efficiency of *tpk1Δ* deletion strain was reduced to 40% whereas the deletion of *yku80Δ* showed an increase in repair efficiency compared to wild type cells (Fig. 3.2) due to a yet undefined mechanism (Boulton SJ and Jackson SP 1996). One interpretation for this increase is the possibility that Yku80 may bind to blunt ends of DNA and prevent them from rejoining

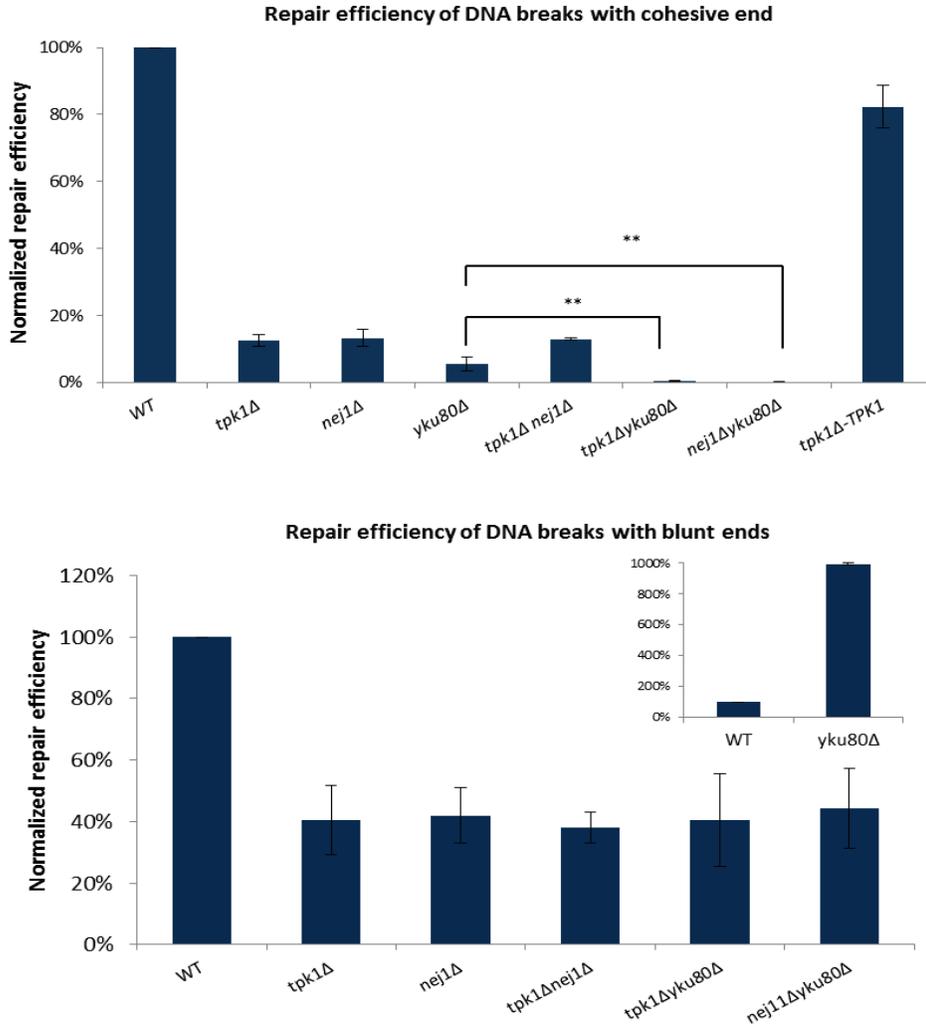


Figure 3.2: Deletion of *TPK1* reduces the efficiency of repair for plasmids with overhangs and blunt ends. In case of breaks with overhangs deletion of *TPK1* (*tpk1Δ*) or *NEJ1* (*nej1Δ*) reduce the repair efficiency to the same level as double gene deletion strain *tpk1Δnej1Δ* while *yku80Δtpk1Δ* and *yku80Δnej1Δ* have a strikingly reduced efficiency in repair. Repair efficiency for blunt end breaks is similar for *tpk1Δ*, *nej1Δ* and *tpk1Δnej1Δ* strains. Indicated in the inset, deletion of *YKU80* (*yku80Δ*) increases the repair efficiency of blunt end breaks. *yku80Δtpk1Δ* and *yku80Δnej1Δ* have reduced efficiencies comparable to those for *tpk1Δ* and *nej1Δ*. The rate of plasmid repair in WT for breaks with cohesive and blunt ends are approximately 75% and 60%, respectively, which are set at 100% for normalization purposes. Error bars represent standard deviation.

via an alternative pathway (Boulton SJ and Jackson SP 1996). To confirm that the observed reductions were due to *TPK1* deletion and not as a result of a possible secondary mutation in our analysis, *TPK1* was reintroduced back into *tpk1Δ* strain. Reintroduction of *TPK1* at the chromosomal level escalated the efficiency of NHEJ to levels comparable to those for a wild type strain (Fig. 3.2). The extent of defects in repair efficiency of *tpk1Δ* mutants which is similar to those for mutants of key NHEJ genes signifies the role of *TPK1* in an efficient repair of DSB.

3.4.5 Efficient repair of chromosomal breaks is dependent on *TPK1*

The JKM139 strain carries an HO-specific endonuclease under the transcriptional control of a GAL promoter. In the presence of galactose, the HO-specific endonuclease is activated causing chromosomal breaks at the MAT locus (Moore JK and Haber JE 1996). HML and HMR, homologous regions to MAT locus, have been deleted in JKM139, making it an ideal system to study NHEJ at the chromosomal level (Shim EY et al. 2005). The survival of a gene deletion mutant strain for a target gene in JKM139 background is compared to that of JKM139 strain when DSBs are induced. In this assay, we subjected the cells to constant DSBs induction for 24 hours. In the event the repair of DSBs at the MAT locus are mutagenic, HO-endonuclease recognition will not be possible therefore the fidelity of repair may play a factor in cell growth. To further study the involvement of *TPK1* in repairing DSBs at the chromosomal level we subjected its gene deletion mutant to a JKM139-based analysis. Equal number of cells for JKM139 gene mutant *tpk1Δ* and control strains were serially diluted and spotted on media containing either glucose or galactose (Fig. 3.3). Survivability was estimated by the number of colonies formed for each dilution. It was observed that when DSB was induced, *tpk1Δ* had a reduced survival rate. This is in accord

with the results obtained using plasmid repair assays, highlighting the significance of *TPK1* in repairing chromosomally induced breaks.

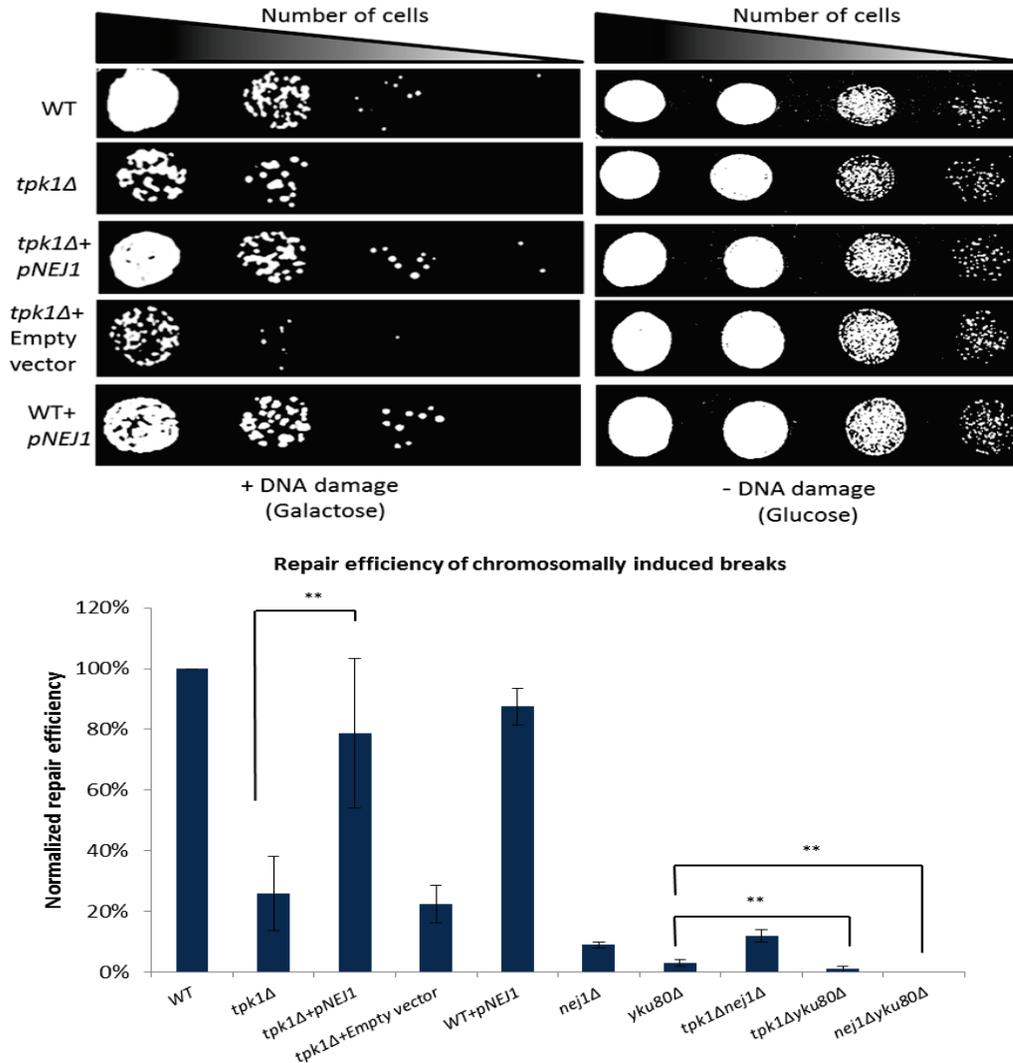


Figure 3.3: Deletion of *TPK1* reduces repair efficiency of chromosomally induced DSB breaks. JKM139 strain contains a site specific HO-endonuclease which is activated in presence of galactose. In this assay, cell survival is used as a measure for repair efficiency. Introduction of *NEJ1* in *tpk1Δ*, reinstates the repair efficiency to levels comparable to those of a control strain. Cell survival is represented as spot tests. Colony counts are shown in a bar graph and represent the average values for at least 5 independent experiments. Error bars represent standard deviation and the p-Values were $7.2E^{-4}$ and $3.6E^{-6}$ for *yku80Δ*, *tpk1Δyku80Δ* and *yku80Δ*, *nej1Δyku80Δ* respectively.

3.4.6 Chromosomal break suppression analysis reveals that *NEJ1* overexpression can compensate the deletion of *TPK1*

TPK1 codes for a kinase known to phosphorylate a diverse group of substrates (Deminoff SJ et al. 2006). Due to the severe phenotype that the deletion of *TPK1* causes in NHEJ efficiency we reasoned that it may also function as a regulator of the activity of a key NHEJ protein. To further investigate the activity of *TPK1* in NHEJ pathway we subjected its gene deletion strain in JKM139 background to suppression analysis using overexpression plasmids. In this way we screened for genes that, when overexpressed, compensate for the absence of *TPK1* in a JKM139 assay. This study is helpful in examining the regulatory effects of different genes within a pathway. Overexpression of a substrate in the absence of a modifier may result in a phenotypic rescue. Additionally, the order that each gene functions within the pathway could be investigated. Plasmids containing *YKU70*, *YKU80*, *MRE11*, *RAD50*, *DNL4*, *LIF1* and *NEJ1* were transformed into *tpk1Δ* deletion strain and wild-type in JKM139 background (Fig. 3.3 and Appendix 3.1). To monitor artificial effects of overexpression plasmids, an empty vector was also transformed into the mentioned strains. In the presence of chromosomal break induction, *NEJ1* overexpression restored *tpk1Δ* cell viability levels back to levels comparable to those for a wild type strain (Fig. 3.3). Overexpression of other key genes of NHEJ in *tpk1Δ* of JKM139 background did not have any significant phenotypic enhancement compared to *tpk1Δ* in JKM139 under the induction of chromosomal break (Appendix 3.1). These results suggest a functional association between *NEJ1* and *TPK1* within the NHEJ pathway. These observations are in agreement with an enzyme-substrate association where overexpression of the substrate could potentially compensate for the absence of the enzyme (Sopko R and Andrews BJ 2008).

3.4.7 Analysis of double mutant strains suggest that *TPK1* and *NEJ1* function in parallel to *YKU80*

To further investigate the functional relationship of *TPK1* and *NEJ1*, plasmid repair assay for *nej1Δtpk1Δ* was carried out. Deletion of two genes that function together in the same pathway often results in a phenotype that is similar to the phenotype of the single deletions. Alternatively, a significant change in the phenotype is expected if the deleted genes work in compensatory parallel pathways. Repair efficiency of *tpk1Δ* is comparable to that of *nej1Δ*, 12.6% and 13.9%, respectively. Double gene deletion strain *nej1Δtpk1Δ* had a repair efficiency of 12.9% which is close to the repair efficiency of single deletions of *nej1Δ* and *tpk1Δ* (Fig. 3.2) further connecting the activities of these two genes in NHEJ.

We also examined the functional relationship between *TPK1* and *YKU80*, a key NHEJ gene. Interestingly, deletion of *TPK1* and *YKU80* (*tpk1Δyku80Δ*) severely impairs the repair mechanism, resulting in an efficiency of 0.03% (Fig. 3.2). Since *yku80Δ* has a low repair efficiency of 5.6%, the inability of *tpk1Δyku80Δ* double deletion strain to repair DSBs suggests that *TPK1* works in parallel to *YKU80*. Double deletion of *YKU80* and *NEJ1*, proposed partner of *TPK1* in NHEJ, reduces the efficiency of repair to 0.17%.

Damage to DNA may result in the formation of DSBs with blunt ends; therefore we further studied the role of *TPK1* and *NEJ1* in repairing DSBs with blunt ends. Blunt end repair through NHEJ is reported to be independent of the activity of YKU complex (Yu L and Volkert MR 2013). Our results indicate that single gene deletions for *TPK1* and *NEJ1* reduces the efficiency of blunt end repair to similar levels, 40% and 42% respectively, whereas the deletion of *yku80Δ* increases the efficiency to approximately 10-fold compared

to wild type (Fig. 3.2) as previously reported, the reason for this increase however, remains unclear (Boulton SJ and Jackson SP 1996). These findings suggest an important involvement for *NEJ1* and *TPK1* in the repair of blunt-ended DNA breaks. As it was evident in the case of DSBs with overhangs, double mutants of *nej1Δtpk1Δ* had no additional effect on the repair efficiency of DSBs with blunt end compared to single mutants of *tpk1Δ* or *nej1Δ*, resulting in a 38% repair efficiency. This further supports our hypothesis that *TPK1* and *NEJ1* function in the same pathway. Our screen also showed that double mutants of *tpk1Δyku80Δ* and *nej1Δyku80Δ* reduce the repair efficiency of blunt end DSBs to 41% and 44%, respectively (Fig. 3.2), which are comparable to repair efficiencies for *tpk1Δ* and *nej1Δ* single mutants. Since phenotypes of *tpk1Δ* mutants and *yku80Δtpk1Δ* double mutants are similar, it's an indication that *TPK1* dictates the behavior of the cell in the repair of blunt ended breaks in a YKU-independent manner. In agreement with our results, previous studies have shown that repair of DSBs with non-cohesive ends is independent of YKU heterodimer (Boulton SJ and Jackson SP 1996). Based on these results, *YKU80* appears not to play a significant role in the repair of DSBs with blunt ends while *TPK1* and *NEJ1* are crucial players in this pathway.

Additionally, we studied the functional connections for *NEJ1*, *TPK1* and *YKU80* in the repair of chromosomally induced breaks. In agreement with our above observation for plasmid repair analysis, *nej1Δtpk1Δ* showed similar repair efficiency compared to *nej1Δ*, supporting a role for *TPK1* in the same pathway as *NEJ1* (Fig. 3.3). In addition, deletion of *TPK1* or *NEJ1* in *yku80Δ* showed additional reduction in the ability of cells to repair chromosomal break repairs further supporting that *TPK1* and *NEJ1* function in a pathway parallel to *YKU80* (Fig. 3.3).

3.4.8 Tpk1 phosphorylates Nej1 in vitro

To further study the Tpk1-Nej1 relationship, a fluorescence-based phosphorylation assay was carried out. Purified proteins were run on SDS-PAGE and stained with coomassie blue to determine the purity of the samples. In this *in vitro* kinase assay, purified proteins are incubated in the presence of ATP and the production of ADP produced through phosphorylation of a substrate by the kinase is measured (Wu ZL 2011). Incubation of Tpk1 with Nej1 resulted in the conversion of ATP to ADP 12-fold higher than the background level measured by the incubation of Nej1 alone, set at 1.00 (Fig. 3.4A). This finding is consistent with a previously reported study that identified Nej1 as a substrate for Tpk1 (Ptacek J et al. 2005) and further supports an Nej1-dependent role for Tpk1 in NHEJ. Domain analysis of Nej1 suggests that Nej1 may be phosphorylated in three possible sites, S297, S298 and S309. Since the purity of proteins used in this analysis was not thoroughly evaluated, the evidence for Nej1 phosphorylation by Tpk1 remains indirect.

3.4.9 Tpk1 is recruited to the site of damage and its deletion reduces the abundance of Nej1 upon DSB induction

Recruitment of Tpk1 to the site of DSB was investigated using CHIP analysis. We FLAG-tagged Tpk1 in JKM background, in which a DSB is created in the presence of galactose through its HO-site-specific endonuclease (Faucher D and Wellinger RJ 2010). Samples were collected before DSB induction and after 30 minutes of induction to measure Tpk1 abundance at the site of damage. Our data showed that the amount of Tpk1 present at the site of damage increased by over ten-fold within 30 minutes of the break induction when

compared to undamaged DNA (Fig. 3.4B), suggesting that Tpk1 is recruited to the DNA break site upon DNA damage.

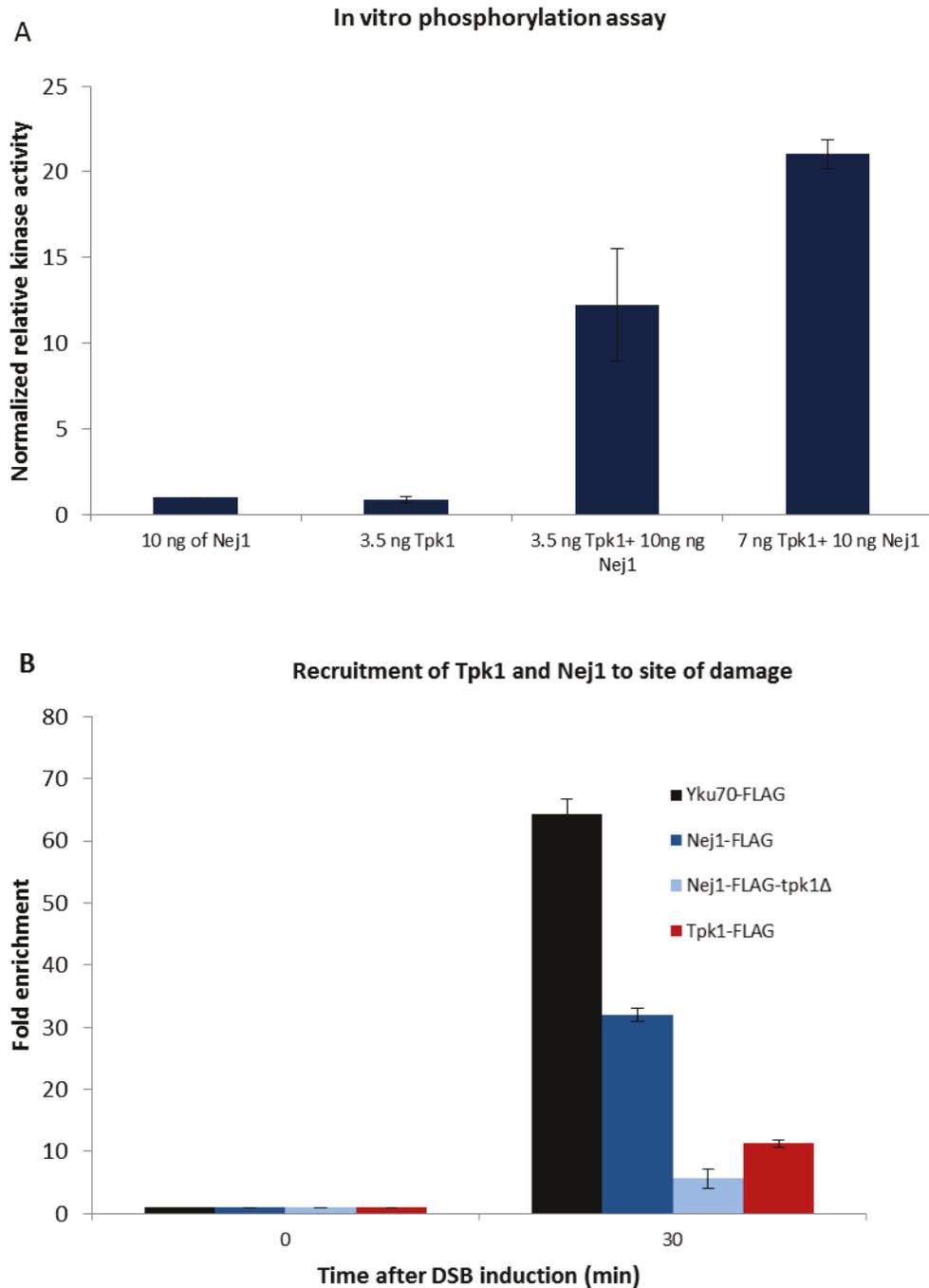


Figure 3.4: A) Kinase activity of Tpk1. The activity of Tpk1 was measured using purified Tpk1 and Nej1 tagged with GST, using fluorescence based *in vitro* assay. Incubation of Tpk1 with Nej1 resulted in ADP production indicating a kinase activity. Error bars represent standard deviation. B) ChIP analysis showed that Tpk1 and Nej1 were recruited to the site of damage after 30 minutes of break induction. Deletion of *TPK1* significantly lowered recruitment of Nej1 to the site of damage. Error bars represent standard deviation.

Since our genetic and phenotypic assays suggested the Tpk1 role in NHEJ is dependent on Nej1, we next examined the recruitment of Nej1 to the site of damage. We FLAG-tagged *NEJ1* in JKM wild type and JKM *tpk1Δ* mutant strains to investigate an association between Nej1 recruitment and the presence of Tpk1 at the site of damage. We observed that the deletion of *TPK1* dramatically reduces recruitment of Nej1 to the site of damage. After 30 minutes of induced damage, Nej1 abundance at the site of damage was reduced by six-fold in *tpk1Δ* strains relative to wild type (Fig. 3.4B). Altogether, these observations further illustrate a functional association between Tpk1 and Nej1 during NHEJ.

3.4.10 *TPK1* role in NHEJ is conserved in mammalian cells

After confirming that *TPK1* functions in NHEJ, we assessed its NHEJ activity *in vitro* in the model human osteosarcoma *U2OS cell line* using *PRKACB*, the human homolog of yeast Tpk1. To study the repair efficiency of NHEJ in mammalian system, we used an established reporter construct (Seluanov A et al. 2004), where the GFP coding sequence is intervened by inserting an adenoviral exon to inactivate the expression of GFP signal. However when the construct is digested with HindIII or I-SceI, the exon is cut out and a DSB is created (Appendix 3.2). Only the successful repair of DSB by NHEJ will restore the GFP expression which can be assessed through FACS by calculating the percentage of GFP expressing cells. Most notably, consistent with *TPK1* in yeast, knockout of *PRKACB* in U2OS cells significantly ($p\text{-value} \leq 1.3 \times 10^{-7}$) reduced the NHEJ-mediated DSB repair as compared to cells transfected with the empty LentiCRISPR vector control (Fig. 3.5), suggesting a conserved functional role of *TPK1* in NHEJ repair pathway. Disruption of *PRKACB* was validated via genomic sequencing (Appendix 3.3).

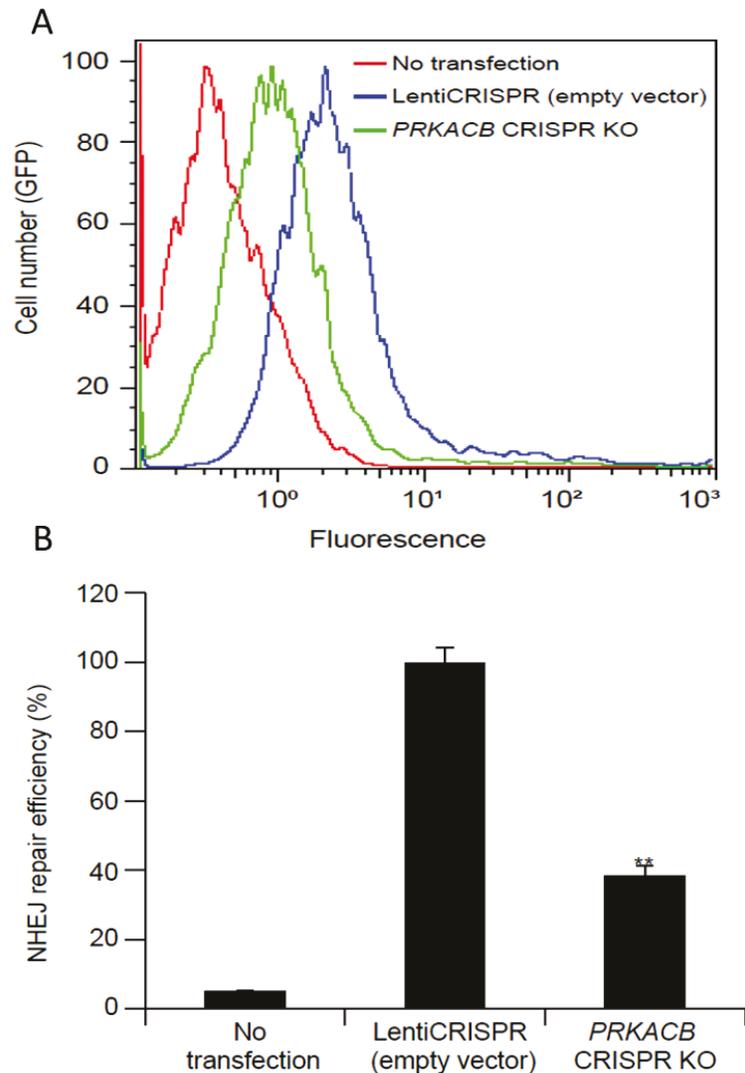


Figure 3.5: *PRKACB* is critical for NHEJ. (A) Histograms showing the number of FACS-sorted positive GFP cells (with pEGFP-Pem1-Ad2 reporter) in U2OS transfected with empty LentiCRISPR vector (blue) or *PRKACB* CRISPR knockout (green). U2OS cells not transfected with any of the indicated strains (red) served as control. (B) Quantitative measurements of NHEJ repair efficiency in the indicated CRISPR knockout and empty vector strains. Error bars show measurement of standard deviation from six independent experiments (p -value $\leq 1.3e^{-7}$).

3.5 Concluding remarks

Kinases are important regulators of protein activity. They participate in regulation of the majority of cellular processes including DNA damage repair. Kinases such as Dun1, Rad53, Tel1 and Mec1 are key regulators of DNA repair pathways, playing important roles in directing the cell to a specific repair pathway and in the recruitment of repair proteins (Branzei D and Foiani M 2006). Here, we report a novel role for Tpk1 in the process of NHEJ through its interaction with Nej1. Tpk1 is one of the key kinases localized to the nucleus (Sopko R and Andrews BJ 2008) that has been previously implicated in the DNA damage response, yet its function in the process is not fully understood (Searle JS and Sanchez Y 2004). Tpk1 carries certain characteristics of repair proteins; being localized to the nucleus, having kinase activity and here we predict PPI interactions for its human homolog with a number of proteins linked to DNA repair. A functional association for *TPK1* with *NEJ1* can explain the severe DSB repair deficiency phenotypes observed when *TPK1* is deleted. XLF, human homolog of Nej1, affects mammalian NHEJ in a similar manner to Nej1 (Callebaut I et al. 2006). Similar to Nej1, XLF C-terminus domain is important to its function in NHEJ, as it is the site of interaction with XRCC4 and is required for the XLF-XRCC4 complex interaction with DNA (Andres SN et al. 2012). Recent studies have shown that phosphorylation of XLF negatively regulates its function in NHEJ in *Schizosaccharomyces pombe* and mammalian cells (Hentges P et al 2014 and Liu P et al. 2015). Additional investigations into the phosphorylation of yeast Nej1 in the context of DNA repair would help us further elucidate the activity of this protein in yeast. Our data further supports a role for PRKACB, the human homolog of Tpk1, in DSBs repair through NHEJ in mammalian system, suggesting a conserved role for Tpk1 in NHEJ between yeast

and human. In the future it will be of interest to investigate the functional association between PRKACB and XLF in a mammalian system.

Chapter 4: Actin Related Protein 6, Arp6, is involved in repair of DNA double stranded break in *Saccharomyces cerevisiae*

4.1 Abstract

DNA damage in form of a double stranded break (DSB) is the most severe form of DNA lesions. These types of damage are generally repaired via non-homologous end joining (NHEJ) through direct ligation or homologous recombination (HR) using a template sequence. NHEJ repair requires three main complexes: Yku70/Yku80 (KU) to protect DNA ends, Mre11, Rad50, and Xrs2 (MRX) for end bridging and Dnl4, Lif1 and Nej1 for ligation. Initiation and regulation of DSB repair requires chromatin remodeling complexes, INO80 which is associated with HR, SWR1 which regulates NHEJ and RSC which is involved in both repair pathways and DNA damage response. Here we identify, Arp6 as a novel factor involved in NHEJ and potentially HR. Our data indicates deletion of *ARP6* reduces repair efficiency of breaks with cohesive ends and chromosomally induced breaks. Furthermore, deletion of *ARP6* enhances fidelity of repair resulting in more accurate rejoining compared to WT strain. Repair efficiency of *arp6Δrsc2Δ* and *arp6Δmre11Δ* double mutants showed no additional reduction compared to single mutants. Contrarily, further reduction in repair efficiency was observed in double mutants of *ARP6* and members of INO80 and SWR1 complex compared to single mutants suggesting that *ARP6* functions independently of SWR1 and INO80 complexes. We further showed that overexpression of *ARP6* rescued the phenotype of *rsc2Δ* and *mre11Δ* under DNA damaging conditions. These observations implicate *ARP6* in the same pathway as *RSC2* and *MRE11*. We also show *arp6Δ* mutants cells result in defective HR repair.

4.2.1 Introduction

The integrity and stability of the DNA is constantly threatened by exogenous and endogenous sources damaging the DNA in different forms. DNA double strand breaks (DSBs) are the most genotoxic form of DNA lesions resulting from breaks in the phosphodiester backbone of both DNA strands (Schipler A, Iliakis G 2013). These forms of breaks arise from various sources including ionizing radiation, reactive oxygen species (ROS) and replication stress. Mis-repaired and unrepaired DSBs can cause genomic instability, chromosomal rearrangement and may lead to carcinogenesis or cell lethality (Chapman J.R et al 2012). DSB repair is achieved by two main distinct pathways: homologous recombination (HR) and non-homologous end joining (NHEJ). Repair through HR is considered error-free due to use of a homologous sequence as template. NHEJ repairs DSBs by re-joining the broken ends of DNA resulting in an error-prone process (Shrivastav M et al 2008). NHEJ is highly conserved from single cell organisms such as yeast, *Saccharomyces cerevisiae* to human. NHEJ plays a primary role in DSBs repair in mammalian cells, though it is a secondary mechanism of repair in *S. cerevisiae* (Hefferin ML and Tomkinson AM 2004).

NHEJ repair in *S. cerevisiae* requires three main protein complexes for initiation of repair, formation of bridge over broken ends and end joining steps. In the initial step, heterodimer of Yku70/Yku80 form a ring over each broken end of the DNA, stabilizing it and protecting it from degradation. Next the MRX complex (consisting of Mre11, Rad50 and Xrs2) bring the DNA ends in close proximity. Finally Dnl4 and Lif1 perform ligation by filling the gaps and joining the broken ends finalizing the repair process (Davis DJ and Chen AJ 2013). Recruitment of the repair proteins to the site of damage requires a cross-talk

between DNA damage response proteins, chromatin remodeling complexes and histone modifications (Sinha M and Peterson CL 2009). In yeast Mec1 and Tel1, orthologs of human ATR and ATM, are recruited to the break site through a series of signaling kinases. These two key DNA checkpoint proteins phosphorylate H2A in response to DSBs, which results in recruitment of histone modifiers and chromatin remodeling complexes (Attikum HV and Gassar SM 2009). A key protein in this process is Arp4, which interacts with phosphorylated H2A and recruits chromatin remodeling complexes (Downs JA et al. 2004). Repair proteins access the DNA through changes in chromatin structure by chromatin remodeling complexes and histone modifications.

In response to DSBs in yeast, chromatin remodeling complexes INO80, RSC and SWR1 are recruited to the site of damage (Price BD and D'Andrea AD 2014). INO80 and SWR1 are recruited to the site of damage in an Arp4-dependant manner, a common protein involved in both complexes (Downs JA et al. 2004). These complexes have been shown to facilitate exchange of H2A with its variant H2A.Z during DSBs repair (Price BD and D'Andrea AD 2014). INO80 plays a predominant role in repair of DSBs at the replication forks and promotes end resection in HR dependant repair. On the other hand, SWR1 facilitates recruitment and binding of Yku complex to the site of break in preparation for NHEJ repair (Attikum H et al. 2007a). RSC plays various roles in DSBs repair, as its members are reported to have distinct functions in different repair pathways and DNA damage response (DDR) (Liang B et al 2007). H2A phosphorylation level after damage induction is dependent on RSC (Shim EY et al. 2007). Additionally members of RSC complexes, Rsc2 and Rsc1, are reported to interact with Yku80 and Mre11 and function in NHEJ repair (Chambers AL et al 2012).

Highly conserved members of Actin Related Protein (ARP) family are common to all three chromatin remodeling complexes involved in the repair of DSBs. RSC complex contains two ARP members, Arp9 and Arp7, INO80 has three members of ARP proteins, Arp8, Arp5 and Arp4 which is also part of SWR1 complex alongside Arp6 (Szerlong H et al. 2003, Tosi A et al 2013 and Chambers AL 2012). These nuclear proteins play important roles in chromatin remodeling, transcription regulation and DNA damage response. Arp6 is reported to interact with Swc2, Swc6 and Swc3 and perhaps forming a sub-complex within SWR1 complex (Wu WH et al. 2005). Arp6 function in regulation of gene expression is tied to its role in exchanging H2A with the variant H2A.Z in an SWR1-dependant manner. Additionally, Arp6 is reported to play an additional role in regulating ribosomal proteins in an Swr1 independent manner (Yoshida T et al. 2010). Here we report a novel function for Arp6 in DSBs repair through NHEJ independent of SWR1 complex. Our genetic analysis suggests that *ARP6* and *RSC2* function in the same pathway during NHEJ repair, and Arp6 role is independent of SWR1 complex and Htz1. In addition, we show that deletion of Arp6 reduces the efficiency of HR pathway.

4.3 Material and Methods

4.3.1 Yeast strains and plasmids:

Deletion strains used in this study were constructed in BY4741 background (MATa orf Δ ::KanMX4 his3 Δ leu2 Δ met15 Δ ura3 Δ) (Winzeler et al. 1998). For SGA analysis *ARP6* was deleted in Y7092 (MAT α can1 Δ ::STE2pr-HIS3 lyp11 Δ ura31 Δ leu21 Δ his31 Δ met151 Δ) background by PCR transformation method as in (Tong AH and Boone C 2007), target gene was replaced by NatR. Deletion of *ARP6* in JKM139 strain (MATa hmr Δ ::ADE1

hml Δ ::ADE1 ade1-100 leu2-3,112 lys5 trp1::hisG ura3-52 ade3::GAL-HO) was constructed through PCR transformation, in which *ARP6* was replaced by Nat^R. Plasmid repair assays were carried out using plasmid p416 containing URA marker and Amp^r selection. Blunt end repair assay was carried out with YCplac111 with LEU marker and Amp^r gene. Plasmid pMV32 carrying Leu2 marker and Amp^r was used to study repair accuracy of *ARP6* mutants (Bahmed K et al. 2010). Different gene expression plasmids used in suppression analysis are derivatives of pGEH plasmid, which contains URA marker, Amp^r selection and GAL promoter linked to each gene.

4.3.2 Protein-protein interaction prediction

Prediction of protein-protein interactions were performed as described in (Pitre S et al. 2012 and Chapter 2). In this method short co-occurring conserved polypeptide regions of reported interacting proteins are used to predict novel interactions. A list of reported (known) interactions were obtained from (BioGRID: www.thebiogrid.org and DIP: www.dip.doe-mbi.ucla.edu).

4.3.3 Plasmid repair assay

Intact plasmids Ycplac111 and p416 were digested at SmaI and XbaI site, respectively. These sites lacked homology to chromosomal DNA. Parallel transformation of digested and circular plasmid was carried out in WT and mutant strains. After 2-3 days of growth on synthetic minimal media lacking uracil or leucine, colonies were counted (Omid K et al. 2014). Ratio of colonies formed by digested transformants to circular transformants in mutant strains was normalized to WT strain indicating the repair efficiency.

4.3.4 Repair fidelity assay

Plasmid pMV1328 was digested by PstI within its KanMX6 marker. Digested and circular plasmids were transformed into mutants and WT strains of JKM139 background. Cells were grown on synthetic medium lacking leucine for 2-3 days. To test the accuracy of repair 50-100 colonies from transformations with digested plasmids were grown on YPD medium containing G418. Growth was examined visually and number of viable cells represented accurate repair (Bahmed K et al. 2010). To calculate repair fidelity the ratio of viable cells to total colonies of mutant strains were normalized to the same ratio of WT.

4.3.5 Drug sensitivity tests

Mutants and WT strains were grown to saturation in YPD medium for 2 days. Subsequently, serial dilutions of each strain were prepared, and 15ul of each dilution was spotted on YPD media containing 3 ul/ml phleomycin or 70 mM hydroxyurea (HU) and drug free medium as control. Cell viability of each strain was examined after 2-3 days of growth at 30° C.

4.3.6 Synthetic genetic array analysis

Synthetic genetic array analysis (SGA) was performed as in Tong and Boone (2007), to map the genetic interaction of *ARP6*. Strains carrying deletion or overexpression of *ARP6* in α -mating type were crossed to DNA damage array and a control array. DNA damage array contains deletion of genes functioning in DNA repair, DNA damage response, cell cycle and chromatin remodeling. Control array is a collection of genes with random functions in various processes not related to DNA damage repair. After series of selections the resulted progeny will either carry a double gene mutant or a single gene mutant with *ARP6*

expression. The resulting strains were then grown on medium containing HU and PHLM to test conditional genetic interaction. Final results were analyzed using SGAtools software. Each experiment was repeated three times and the average values are presented.

4.3.7 Chromosomal repair assay

JKM139 strain was used to examine efficiency of repair of chromosomally induced breaks. In this strain, HO-site specific endonuclease has a GAL promoter. In the presence of galactose, the endonuclease is activated producing DSBs at the MATa locus. HML and HMR, homologous regions of MATa, are deleted to prevent the possibility of HR repair (Jessulat M et al. 2008). WT, mutants and mutants carrying overexpression plasmids were grown in YPD or selective media to saturation and were then serially diluted to concentrations of 10^{-1} to 10^{-4} . 15 μ l of each dilution was spotted on YP-Galactose and YPD. Cell viability of mutant strains was compared to that of WT after 2-3 days.

4.3.8 Homologous recombination assay

In recombination assay, plasmid pGV-256-dead was digested at its non-functional *LacZ* gene with BglII restriction enzyme. PCR product containing functional *LacZ* was obtained using pGV-256-live plasmid as template. Linearized plasmid and PCR products were co-transformed into mutant and WT strains and grown for 2 to 3 days (Jessulat M et al. 2008). Minimum of 50 colonies were transferred to a new plate and grown for 1 day. To measure recombination repair β -galactosidase lift assay was carried out on transferred colonies. To calculate recombination efficiency, the ratio of blue to white colonies of mutant strain was normalized to the ratio of blue to white colonies of WT.

4.4 Results

4.4.1 PIPE interactions of ACTL6B homolog of Arp6

Generally, interacting proteins are thought to be functionally related. Therefore a growing number of studies have used physical associations between proteins to identify protein functions (Krogan NJ et al. 2006 and Butland G et al. 2005). Computational tools have been utilized to predict protein-protein interaction (PPI) network with high efficiency and accuracy at a relatively low cost (Pitre S et al. 2008 and Muley VY and Ranjan A 2013). Recently a computational method called Protein Interaction Prediction Engine (PIPE) was used to predict human interactome based on short co-occurring primary protein sequences (Pitre S et al 2012, and Schoenrock A et al 2014). On the basis of the same principles, here we predicted the human protein ACTL6B to interact with several proteins involved in DNA repair (Fig. 4.1). One of ACTL6B interacting partner is CHD4, a chromatin remodeling protein which is involved in DNA damage response and regulation of DSBs repair (Larsen DH et al 2010 and Pan MR et al. 2012). ACTL6B was also predicted to interact with STK4 and STK3, part of MST kinase family (Thompson BJ and Sahai E 2013). STK3 (MST2) is activated in response to DNA damage implicating a role for it in an ATM-dependent pathway (Garth Hamilton et al. 2009). In response to DNA damage, ATM and ATR are key kinases that are activated and initiate a signaling cascade in DDR pathway. ATM is generally activated in response to DSBs and phosphorylates downstream proteins such as Chk1 and p53 (Maréchal A and Zou L 2013). SKT4 (MST1) regulates DNA damage checkpoints in G1 phase, and upregulates p53 expression promoting apoptosis in response to DNA damage (Yuan F et al 2011 and Richardson E et al. 2012). Based on ACTL6B interactions with

proteins involved in DNA damage and repair we hypothesize it plays a role in DNA repair. A BLAST searched

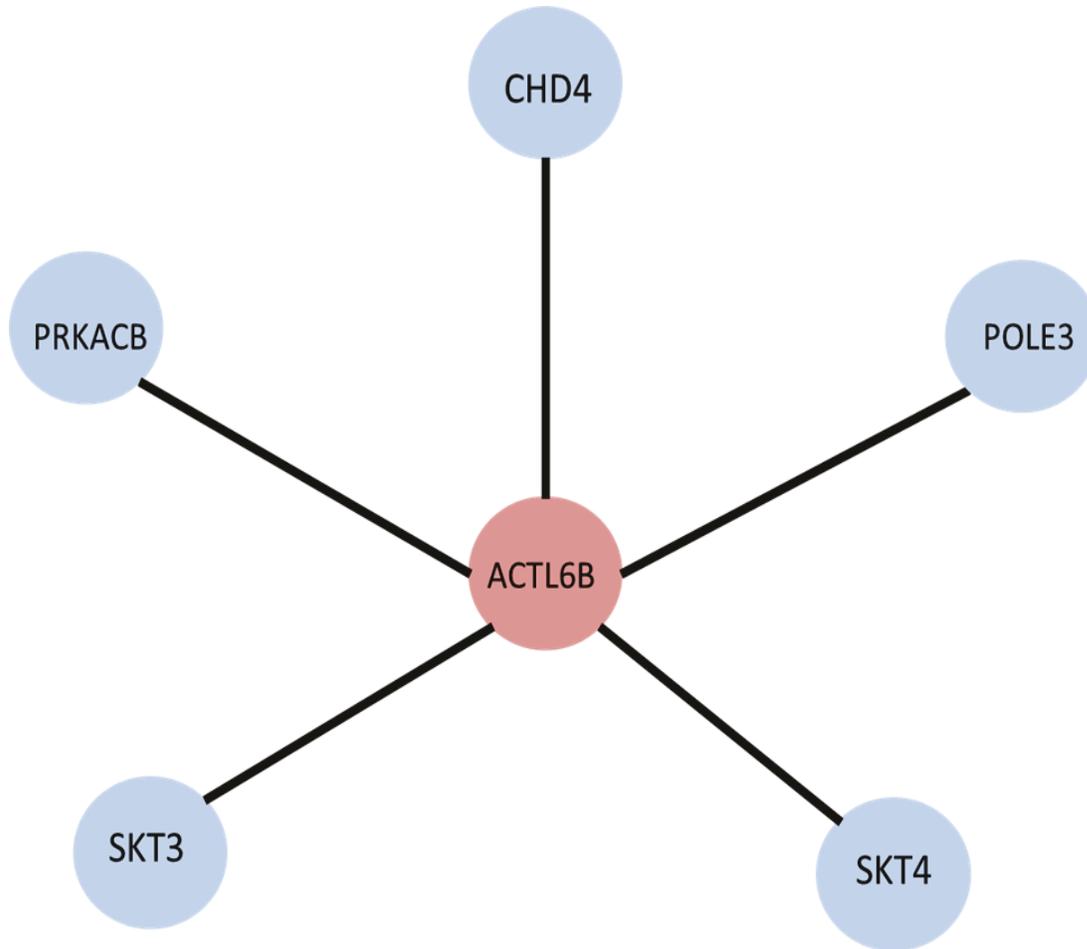


Figure 4.1: PIPE predicted five novel interactions for ACTL6B, which shares 25% identity with Arp6 in yeast. CHD4, SKT3 and SKT4 have been reported to function in DNA repair and DDR.

identified *S. cerevisiae* Arp6 to have significant homology (25.1% identity with a p-value of $19E^{-19}$) with human ACTL6B, therefore we investigated the role of Arp6 in DSBs repair in yeast.

4.4.2 *ARP6* mutants have reduced efficiency of repair in a plasmid based assay

One of the classic assays to investigate NHEJ genes is examining the ability of the gene deletion mutants in rejoining linearized plasmids. Based on this assay, deficiency in ligation of digested plasmid by gene mutant is an indication of involvement of the gene in the repair pathway (Jessulat M et al. 2008). An advantage of this assay is the ability to examine both types of break ends, overhangs and blunt ends (Bahmed K et al 2011). Basically plasmids are digested at a unique site which has no homology to chromosomal DNA. Linearized and circular plasmids are then independently transformed into mutant and WT strains. Finally, repair efficiency is calculated as the ratio of colonies formed by linearized plasmid to colonies formed by circular plasmid of mutant and normalized to that of WT. In case of breaks with overhangs, repair efficiency of DNA breaks was reduced to approximately 40% in *ARP6* deletion strain, while efficiency of repair for positive controls *mre11Δ* and *yku70Δ* were 21% and 8% respectively (Fig. 4.2). We then examined efficiency of *arp6Δ* mutant strains in repair of blunt ended breaks. Of interest deletion of *ARP6* enhanced repair of blunt ended breaks compared to WT resulting in efficiency of 160% (Fig. 4.3A). Similar observation has been made with deletion mutant strains for *YKU70* (Yu L and Volkert MR 2013 and Bahmed K et al 2010) but not for other key NHEJ genes. Based on these data, *ARP6* plays a role in repair of breaks with cohesive ends, while its deletion promotes repair of breaks with no overhangs.

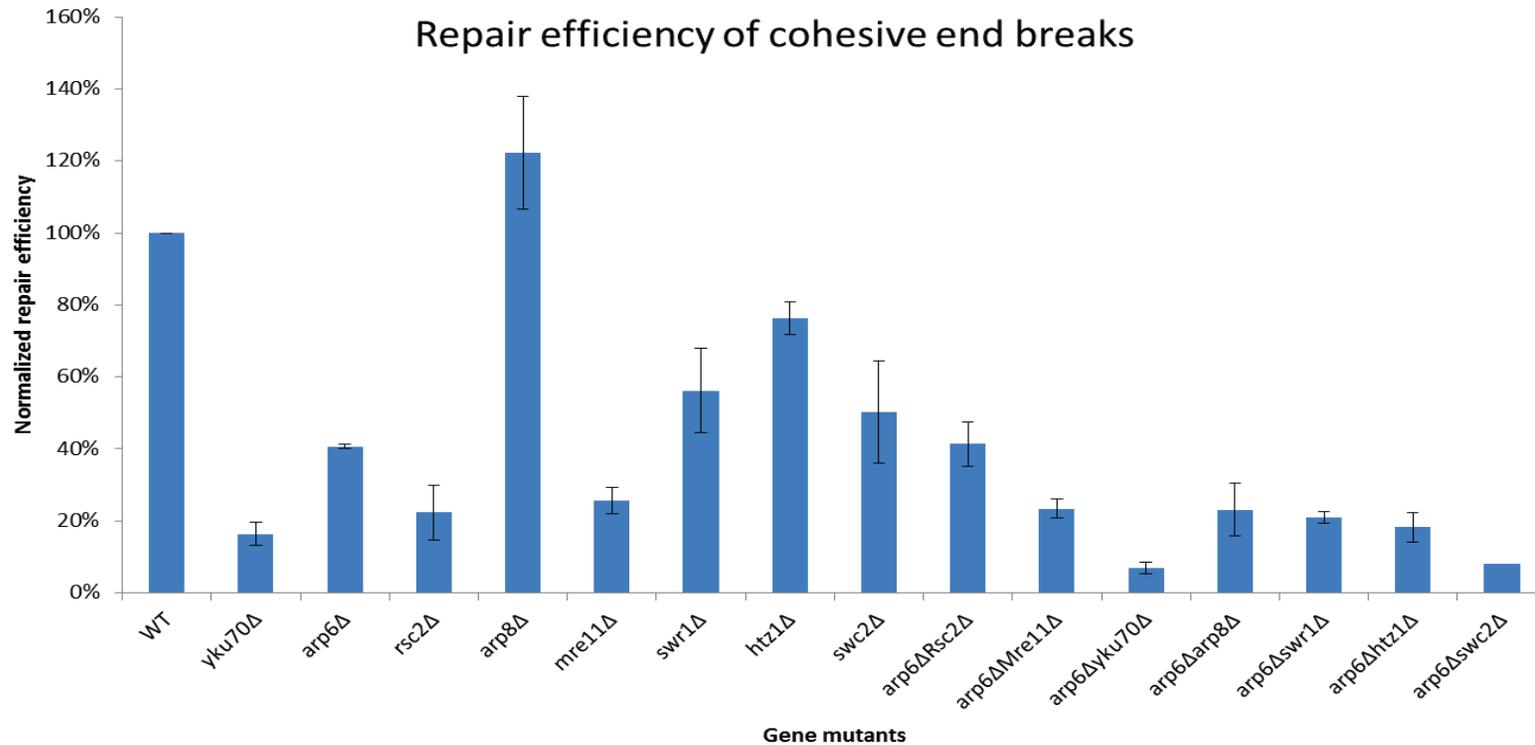


Figure 4.2: Deletion of *ARP6* reduced efficiency of repair of breaks with cohesive ends. Interestingly double deletion of *ARP6* with each of *RSC2* and *MRE11* did not result in additive phenotype, suggesting *ARP6* works in the same pathway as *RSC2* and *MRE11*. Each experiment was repeated at least 5 times.

4.4.3 Deletion of *ARP6* promotes error-free repair by NHEJ

NHEJ is considered an error-prone repair pathway, introducing mutations within the repaired region of the DNA (Chiruvella KK et al. 2013). We examined the role of *ARP6* in carrying out accurate repair by subjecting *arp6Δ* to repair fidelity assay. In this assay, pMV1328 plasmid digested within the KanMX6 gene is transformed into WT and mutant strains. Colonies obtained from transformation are grown on medium containing G418 which requires functional KanMX6 for cell growth. In case of accurate repair, KanMX6 is functional allowing cells to grow on G418 media, while errors during repairs will result in defective KanMX6 gene therefore inhibiting cell growth on medium containing G418 (Yu L and Volkert MR 2013). In this system, accuracy of repair in *arp6Δ* mutants was measured at 216% compared to WT, though the efficiency of repair of *arp6Δ* was 18% (Fig. 4.3B). This suggests that even though deletion of *ARP6* causes defects in the repair process, it carries out the repair with very high accuracy. This data suggests that *ARP6* may affect the balance between efficiency and accuracy of repair.

4.4.4 *Arp6* is involved in repair of chromosomally induced breaks

The chromosomal break assay indicates involvement of genes in repair of breaks induced on a chromosomal level. JKM139 strains have a site specific HO-endonuclease linked to a GAL promoter, in the presence of galactose the endonuclease induces DSBs at the MAT locus (Lee SE et al. 1998). In this strain homologous regions to the MAT locus, HML and HMR, are deleted therefore limiting repair of DSBs to NHEJ pathway. To examine involvement *ARP6* in NHEJ, viability of cells carrying *ARP6* deletion in JKM139 is

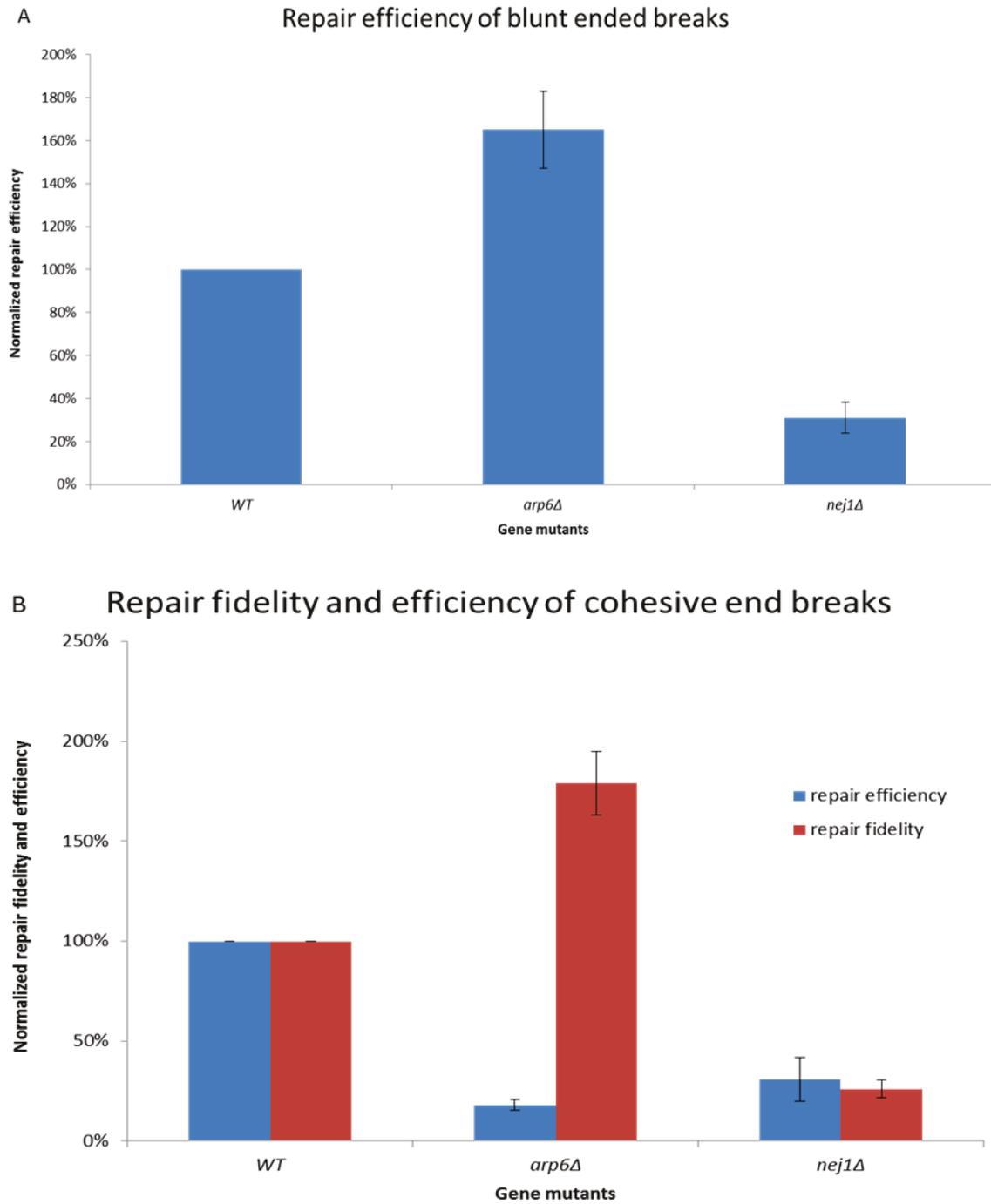


Figure 4.3: A) Deletion of *ARP6* results in enhanced repair of breaks with blunt ends
 B) Accuracy of repair is increased in *arp6Δ* mutant strains. Each experiments were carried out in replicates of 5.

compared to WT under the induction of chromosomal break. Equal amount of *arp6Δ* and WT were serially diluted and spotted on media containing galactose and glucose. Our data showed severe reduction in the survivability of *arp6Δ* compared to control strain under DSBs induction (Fig. 4.4). These results show *ARP6* is significant in repair of chromosomally induced breaks supporting its involvement in NHEJ. These observations are also in agreement with the findings above where *ARP6* deletion increased the fidelity of DSB repair. In chromosomal break assays, the fidelity of repairs play an important role since alterations in the DSB region after the repair would hinder the cell resistant to further assault by HO-endonuclease. Consequently, extreme sensitivity observed for *ARP6* deletion mutant strains can be explained by an increase in the fidelity of the repair in chromosomal break assay.

4.4.5 *ARP6* deletion is sensitive to DNA damaging drugs

Generally, deletion of DNA repair genes increases vulnerability of the cell to DNA damaging drugs (Cagney G et al. 2007). These drugs randomly target various locations on the chromosomes generating different types of DNA lesions. Damages induced by these drugs are not exclusively in the form of DSBs, rather these lesions include SSBs, DSBs, and base and nucleotides damages (Brown JA et al. 2006). Efficient and effective response to DNA lesions caused by DNA damaging agents requires a range of DNA repair genes. Therefore sensitivity of gene deletion mutants to DNA damaging drugs is used as an indicator for the involvement of the targeted genes to DNA repair pathways (Jelinsky JA et al. 2000 and Chang et al. 2002). HU and PHLM are DNA damaging drugs that induce breaks at the chromosomal level by different mechanisms. HU depletes dNTPs pool resulting in the collapse of replication machinery producing SSBs and DSBs (Koç A et al. 2004). PHLM

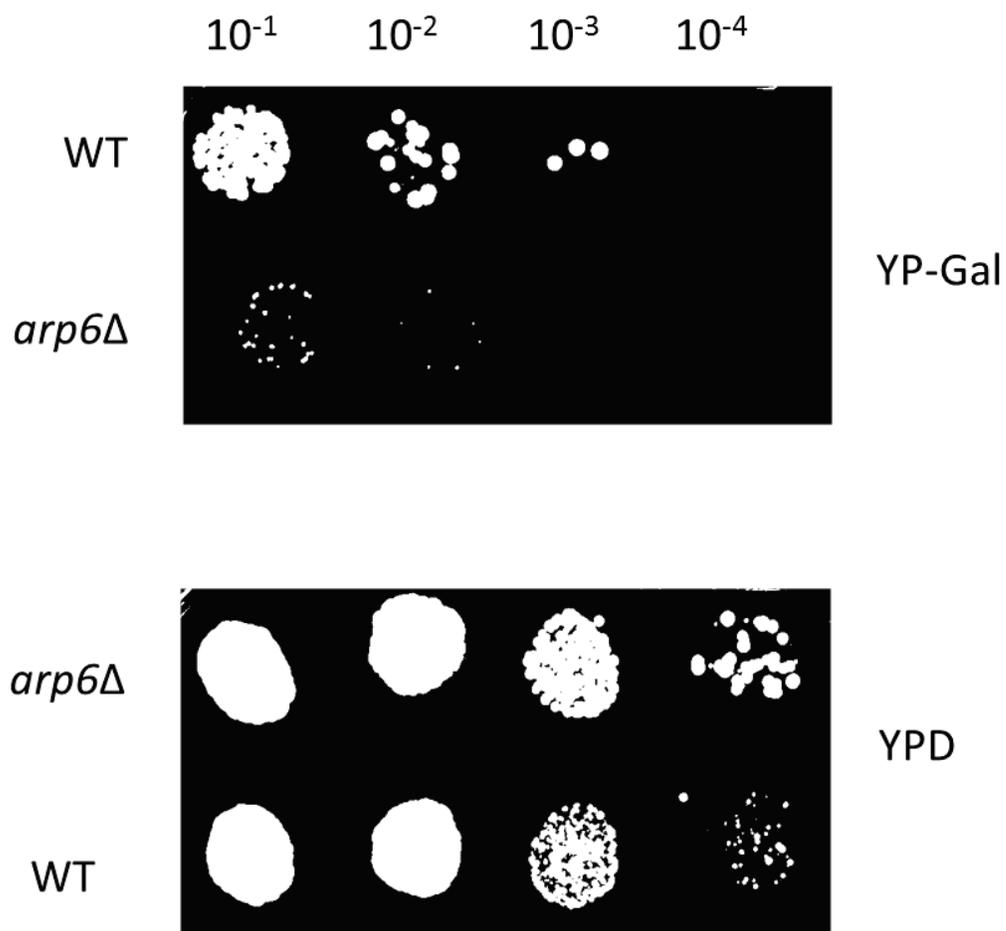


Figure 4.4: JKM 139 contains a HO-site specific endonuclease linked to GAL promoter, inducing breaks at the chromosomal level presence of galactose. Cells carrying *ARP6* deletion show high sensitivity to chromosomally induced breaks.

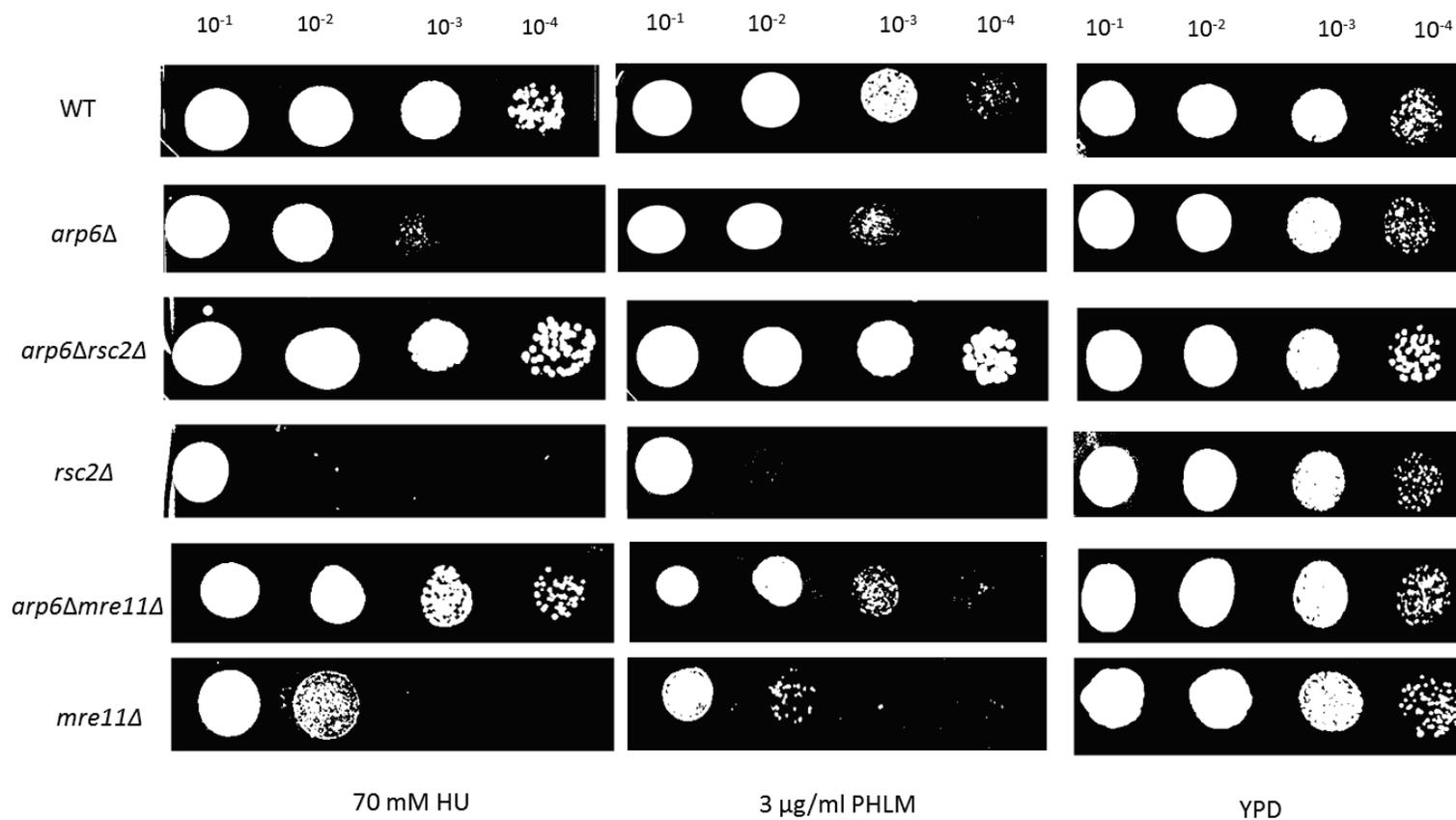


Figure 4.5: Viability of *arp6Δ* mutant was reduced in the presence of DNA damaging agents, PHLM and HU. Conditional positive genetic interaction was observed between *ARP6* and *RSC2* as well as *MRE11*. Phenotypic enhancement is evident in *arp6Δrsc2Δ* and *arp6Δmre11Δ* compared to single

promotes production of free radicals that bind to DNA generating various forms of DNA lesions including formation of DSBs (Pvirk LF et al. 1981). Cells with *arp6Δ* were moderately sensitive to PHLM (3 μg/ml) and HU (70 mM) (Fig. 4.5), further supporting a function for *ARP6* in DNA repair.

4.4.6 *ARP6* genetically interacts with several genes involved HR, NHEJ and cell cycle

Genetic Interaction (GI) network of a gene examines its functional association with genes in different pathways (Costanzo M et al. 2011). It is used to elucidate higher level functional links between closely associate pathways. In functional genomics, GIs are used to investigate novel functions or to elucidate the role of genes in various cellular pathways (Babu M et al. 2014 and Totsi E et al. 2014). Furthermore, conditional GIs reveals association of genes under specific stress conditions (Guénolé A et al. 2013). A negative interaction resulting from deletion of two genes can be identified by a sick or lethal phenotype (Boone C et al. 2007). These interactions suggest that the target genes function in compensating and often parallel pathways. To investigate the role of *ARP6* in DNA repair, we mapped its GIs with genes implicated in DNA repair, DNA damage response and other related pathways. Additionally we examined genetic interactions of *ARP6* under the induction of DNA damage by HU. Deletion of *ARP6* was crossed with two arrays containing 386 gene deletions; first array included genes involved in DNA repair, DNA damage response and related pathways, that we termed DNA damage array (DDA), and second array contained genes with random functions as a control. Our SGA analysis showed that *ARP6* genetically interacts with NHEJ genes *RAD50*, *RTT109* and *RAD27*. In addition to NHEJ genes, *ARP6* interacts with key genes involved in DNA damage response (*ARP8*, *RAD53* and *DDC1*), mismatch repair (*MLH1* and *MLH2*), post-replication repair (*RAD18*). Interestingly,

under DNA damaging conditions, negative genetic interactions were observed between *ARP6* and key HR genes *RAD55* and *RAD52* (Fig. 4.6A). Key NHEJ genes generally interact with HR genes and similarly HR genes with those of NHEJ (Koh JLY et al. 2010). The GI network of *ARP6* which included genes involved in DSBs repair pathways, HR and NHEJ, and DNA damage response, hints to a potential role for *ARP6* which might be upstream of repair pathways. This is in accord with the reported function of *ARP6* as a member of chromatin remodeling complex and connects its activity to DNA damage repair.

Phenotypic suppression analysis examines effects of overexpression of an array of genes on the phenotype of a gene deletion strain, under stress conditions (Alamgir M et al. 2010). Phenotypic rescue of a gene deletion phenotype by overexpression of another gene is often observed in genes which carry similar functions in the pathways (Kuzmin E et al. 2014). For example the overexpression of *CHK1* was previously shown to compensate to the phenotypic consequence of *PPH3* deletion under DNA damaging conditions (Omidi K et al. 2014). To this end, we introduced *ARP6* expression plasmids into DDA and random array strains to generate single gene deletion mutant strains that have *ARP6* overexpressed. Under DNA damage induction through *PHLM* and *HU*, *ARP6* overexpression compensated for sick phenotypes of several interesting gene deletion strains (Fig. 4.6B). These genes included, *MMS22*, which is involved in stabilizing the replication fork, *MRE11* which is a key NHEJ and DNA damage response gene, as well as *RSC2* that codes for a component of RSC complex involved in DNA damage response (Bing Liang et al. 2007). *RSC2* is required for efficient repair of chromosomally induced breaks via NHEJ and physically interacts with *MRE11* (Chambers AL et al. 2012). These results further implicate *ARP6* in the repair pathway upstream of DSB repair and suggest a possible connection between them.

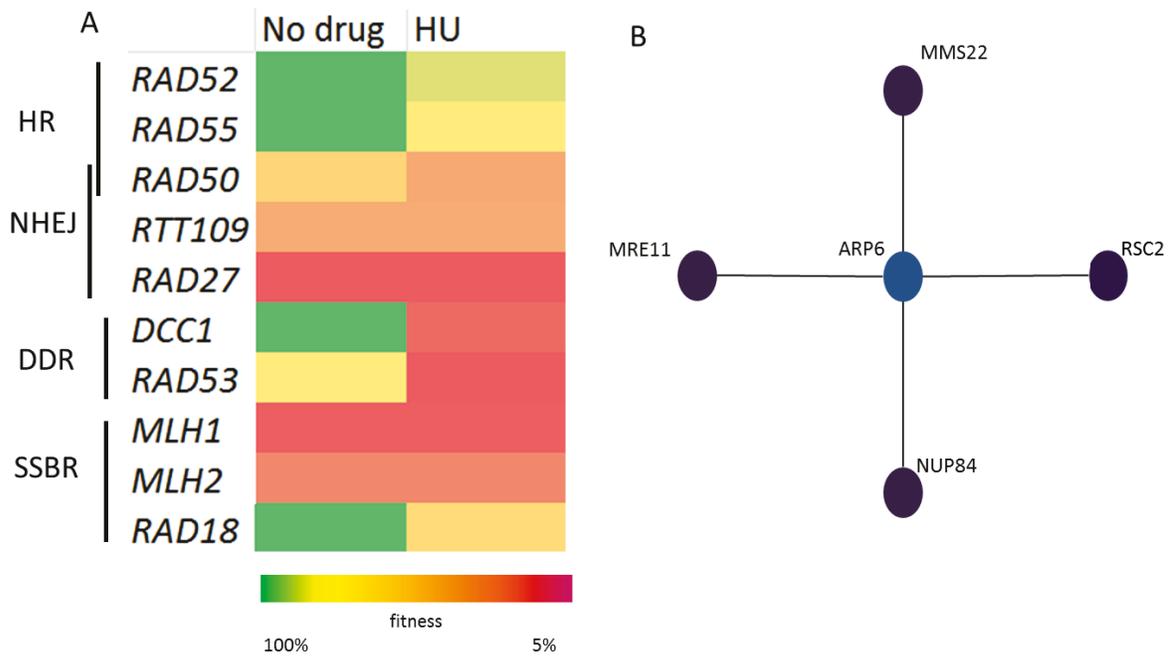


Figure 4.6: A) genetic interaction map of *ARP6* included genes involved in NHEJ, HR and DDR. Interestingly HR genes interact with *ARP6* under DNA damaging conditions. B) Upon induction of DNA damage *ARP6* overexpression compensates for *MRE11* and *RSC2* which are involved in DDR, NHEJ and HR.

4.4.7 Role of *ARP6* in NHEJ is related to RSC complex and independent of SWR1 and SWC

Arp6 is thought to be part of SWR1 complex, which is required for initiation of DSB repair due to its chromatin modifying activity. Even though SWR1 complex is required for response to DNA damage and efficient binding of Yku80, there is no reported role for Arp6 in these processes and its molecular function is not well understood. Other chromatin remodeling complexes that are involved in DSBs repair are INO80, which promotes HR, and RSC which is involved in both HR and NHEJ. Another key protein involved in regulation of DSBs repair is Htz1, a variant of H2A, which is incorporated at the site of break and removed by SWR1 and INO80 respectively (Horigome C et al. 2014). To study the role of Arp6 in NHEJ, we examined its functional relation with members of SWR1, INO80 and RSC chromatin remodeling complexes in addition to Htz1. This was done using analysis of double gene deletion mutants. Generally, deletion of two genes that are involved in the same pathway results in phenotype similar to deletion of each gene individually. However if the genes affect the same process via different pathways, a combinatory effect might be expected where double deletion causes a more severe effect than the deletion of each gene individually.

Since *ARP6* is reported to be a part of SWR1 complex, we first investigated the functional relation of *ARP6* with two members of SWR1 complex, namely *SWC2* and *SWR1*, in re-joining of linearized plasmids. For this, double deletion mutant strains of *arp6Δswr1Δ* and *arp6Δswc2Δ* and single deletion strains of *arp6Δ*, *swc2Δ* and *swr1Δ* were subjected to plasmid repair assay. Deletion of *SWC2* reduced the repair efficiency to 50% similar to that for *ARP6* deletion mutant strain. However, double deletion of *ARP6* and *SWC2*

(*arp6Δswc2Δ*) significantly impaired the repair process with a combinatory effect showing 8% efficiency (Fig. 4.1). In case of single deletion of SWR1, repair efficiency was calculated at 56%, while double deletion of *ARP6* and *SWR1* (*arp6Δswr1Δ*) further reduced the repair efficiency to 20% (Fig. 4.1). These data showed the additive influence of ARP6 with SWR1 and SWC2 and suggest that APR6 influence on NHEJ is independent of SWR1 complex.

Next we examined a role for *ARP6* in NHEJ through INO80 complex and *HTZ1*. *ARP8* is a key member of INO80 complex that is often used for phenotypic analysis (Attikkum HV et al. 2007a). We investigated repair efficiency of double mutants of *ARP6* with *ARP8* and *HTZ1*. Our results showed that although single deletion of *ARP8* did not reduce the efficiency of repair compared to WT, double deletion of *ARP8* and *ARP6* lowered the repair efficiency to 23%, in comparison to that for *ARP6* single deletion which is at 43% (Fig 4.1). Furthermore, *HTZ1* single mutants showed moderate reduction in repair (71%), though double deletion of *ARP6* and *HTZ1* significantly lowered the repair of linearized plasmid to 18% (Fig. 4.1). These observations suggest that *ARP6* role in NHEJ is independent of INO80 and *HTZ1* pathways.

Finally we investigated the role of *ARP6* in association with RSC chromatin remodeling complex. Under DNA damage, recruitment of key DNA damage response proteins, Tel1, Mec1 and Rad9, is shown to be dependent on Rsc2 protein. Additionally, Rsc2 plays an important role in chromatin remodeling at the site of break after induction of DSBs, and it physically interacts with Mre11 (B Liang et al 2007 and Y Shim 2005). To this end we studied functional relation of *ARP6* with *RSC2* and *MRE11* in repair of linearized plasmid. Repair efficiency of *rsc2Δ* was 36% which is comparable to repair efficiencies of *arp6Δ* and *arp6Δrsc2Δ* at 44% and 41% respectively (Fig 4.1). Furthermore, *mre11Δ* and

arp6Δmre11Δ showed comparable repair efficiencies at 26% and 23% respectively (Fig 4.1). Since deletion of *ARP6* does not seem to have additive effect on the reduced efficiency of NHEJ caused by deletion of *RSC2* or *MRE11*, it seems likely that they reduce NHEJ in the same manner placing the NHEJ activity of *ARP6* in the same pathway as *RSC2* and *MRE11*.

4.4.8 Additional genetic evidence for the association of *ARP6* with *RSC2* and *MRE11*

Additionally we investigated the phenotype of *arp6Δrsc2Δ* mutant in response to DNA damaging drugs. Deletion of *RSC2* shows high sensitivity to HU and PHLM causing severe growth defects, while *arp6Δ* mutants are moderately sensitive to both drugs. Surprisingly double deletion of *ARP6* and *RSC2* in the presence of HU and PHLM resulted in phenotypes similar to WT strain (Fig. 4.5). Furthermore, deletion of *MRE11* showed sensitivity to both drugs, though *arp6Δmre11Δ* double mutants resulted in phenotypic suppression. These observations are in agreement with a positive genetic interaction between a pair of genes that function in the same pathway (Costanzo M et al. 2011). In this manner deletion of one gene relieves the cell requirement for the presence of another gene. These data provided further genetic support for a functional association for *ARP6* with *RSC2* and *MRE11*.

4.4.9 *ARP6* is involved in homologous recombination repair

The GIs map of *ARP6* included several genes involved in HR pathway; therefore we investigated a potential role for *ARP6* in HR through a plasmid based assay. In this assay, plasmids are digested within non-functional *LacZ* gene. Linearized plasmids are co-transformed with PCR products containing functional *LacZ* gene into WT and mutant strains. Cells expressing β-galactosidase indicate plasmids acquiring functional *LacZ* through HR,

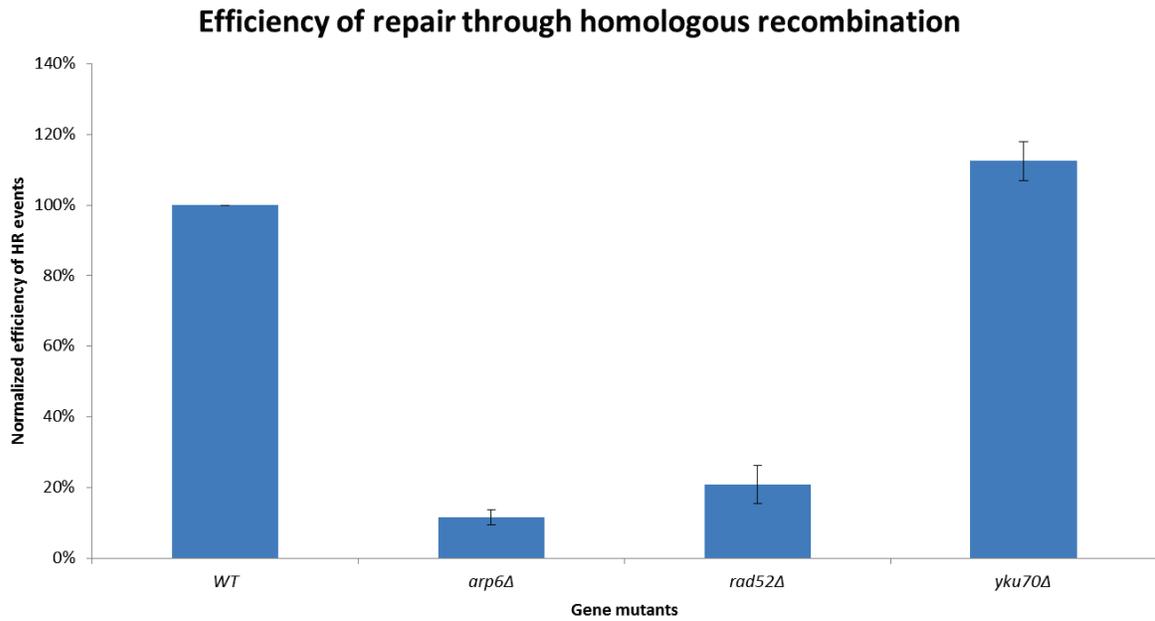


Figure 4.7: deletion of *ARP6* dramatically reduces repair through HR. Cells repairing through HR will acquire functional *LacZ*, in case of *arp6Δ* only 12% of the cells expressed functional *LacZ*. Experiments were repeated 5 times.

while cells carrying plasmids with non-functional *LacZ* repair plasmids through re-joining pathway (Jessulat et al. 2008). Deletion of *ARP6* severely repressed recombination repair, with only 12% of the cells expressing β -galactosidase compared to WT cells (Fig. 4.7). Interestingly efficiency of HR in *arp6 Δ* is lower than positive control *rad52 Δ* in which HR accounts for 24% of repair events. Here we show that *ARP6* functions in repair of DSBs through HR, and its role is not limited to NHEJ. Similar observations have been made for *RSC2* and *MRE11*. In addition to the involvement of *RCS2* and *MRE11* in NHEJ, deletion of these genes is reported to reduce the efficiency of HR.

4.6 Concluding remarks

The actin-related protein family is composed of 10 members including cytoplasmic proteins such as Arp1 and nuclear proteins including Arp4 and Arp6 (Kapoor P and Shen X 2014). In yeast cells, Arp4 is an essential protein which plays an important role in DNA damage response process. Other members of ARP family, Arp8 and Arp5 are also involved in DNA repair process (Morrison A and Shen X 2009). Here we report a novel role for Arp6 in NHEJ linked to RSC complex through its genetic association with Rsc2. Members of chromatin remodeling complexes INO80, SWR1 and RSC are recruited to the site of damage after DSBs induction. SWR1 promotes binding of KU to the DNA ends, and is important in error-free NHEJ repair (Horigome C et al. 2014). Arp6 is part of SWR1 complex, and is involved in regulation of gene expression through exchanging H2A with its variant H2A.Z (Yoshida T et al. 2010). Involvement in SWR1 complex and chromatin remodeling activity of Arp6 makes it an ideal candidate to function in DNA repair process. Efficiency of repair of endogenous chromosomal breaks and chemically induced breaks was reduced in *arp6 Δ* suggesting defect in the repair process in cells lacking *ARP6*.

Genetic analysis showed that *ARP6* role in NHEJ is independent of its association with SWR1 complex. This result is not surprising since *ARP6* has been reported to function independent of SWR1 in other processes such as transcription regulation (Yoshida T et al. 2010). We also reported that *ARP6* role in NHEJ is independent of *ARP8*, another member of ARP family involved in INO80. Studies have shown INO80 role in DSBs repair to be more significant in HR, through recruitment of RPA and Rad51 in an MRX-dependant manner (Tsukuda T et al. 2005). RSC complex has been implicated in DNA damage response pathway, HR and NHEJ (Shim EY et al. 2005). Interestingly *RSC2* is reported to regulate H2A phosphorylation levels in response to DNA damage. Furthermore, Rsc2 is important in recruitment of Tel1 and Mec1 to the site of damage and activation of Rad53 in the DNA damage checkpoint pathway (Liang B et al. 2007). Deletion of *RSC2* reduces efficiency of NHEJ repair further implicating its involvement in DSBs repair pathway, possibly due to its physical interaction with Mre11 and Yku70 (Shim EY et al. 2005 and 2007, Liang B et al. 2007 and Chambers AL et al. 2012). Analysis of the repair efficiency of double gene deletion mutant *arp6Δrsc2Δ* suggested an association between *ARP6* and *RSC2* in NHEJ pathway. Interestingly, double deletion of *ARP6* and *MRE11*, interacting partner of *RSC2*, showed no additional defects in repair, further supporting an *RSC2*-dependent role for *ARP6* in NHEJ. Further biochemical investigations are needed to elucidate the functional association for Arp6 with Rsc2 and Mre11 in the context of DNA repair. It is also interesting to examine whether the role of Arp6 in DNA repair is conserved in other organisms such as human.

Chapter 5: PAS kinase complex regulates the efficiency and fidelity of non-homologous end joining repair

5.1 Abstract

DNA double stranded breaks (DSBs) can be repaired through Non-homologous end joining (NHEJ) and homologous recombination (HR) repair pathways. NHEJ is highly conserved mechanism repairing DSBs through direct ligation of broken ends. This process is initiated by binding of Yku heterodimer to the broken ends, followed by Mre11, Rad50 and Xrs2 (MRX) forming a bridge over the break sites and finally Dnl4, Lif1 and Nej1 ligating the DNA ends. DNA damage checkpoint proteins sense the damage and recruit the key factors involved DNA repair to the site of damage. Mre11 plays a key role in DNA damage response and repair through NHEJ and HR. Another key factor in DDR is Mrc1 which phosphorylates Rad53, a central DNA damage checkpoint protein. Here we report a novel role for the members of PAS kinase complex, Psk1 and Psk2, in DSBs repair via NHEJ. Deletion of *PSK1* and *PSK2* reduces repair efficiency of breaks with cohesive and non-cohesive ends. Additionally, accuracy of repair is significantly reduced in *psk1Δ* and *psk2Δ* mutant strains. We also show that deletions of *PSK1* and *PSK2* make the cells sensitive to DNA damaging agents, a characteristic of DNA repair genes. Our genetic interaction (GI) analysis associates *PSK2* and *PSK1* with DNA damage checkpoint genes *MRC1* and *MRE11*, which are also involved in NHEJ. We hypothesize that role of Psk1 and Psk2 in NHEJ is possibly connected to DNA damage checkpoint activity and DNA damage response process.

5.2 Introduction

DNA molecules are constantly assaulted by both exogenous and endogenous chemicals as well as ionizing radiation. These attacks result in several classes of DNA damage with varying severity, the most severe being the double strand breaks (DSBs). The potential gross-mutagenic or cytotoxic consequences of DSBs invoke an extensive global cellular response. Most notably, there is an intrinsic relationship between the DSB repair and the pathways which facilitate efficient progression through the cell cycle. In response to DSBs, DNA damage checkpoint proteins temporarily halt the cell cycle to provide adequate time to complete repair via different repair pathways. There are two primary independent pathways responsible for the resolution of DSBs, the template-based high fidelity homologous recombination (HR) and the inherently error-prone non-homologous end-joining (NHEJ). HR aligns the break-site with a homologous region of template DNA following end resection, forming dual Holliday junctions (dHJ) and ultimately resolution. The requirement for a homologous template makes HR insignificant in repair of DSBs during G1 phase (Chapman JR et al. 2012). Repair via NHEJ occurs during all phases of the cell cycle but is mainly preferred in G1. During this process broken ends are aligned in close proximity, processed if necessary, and directly ligated by DNA ligase IV (Chiruvella KK et al. 2014). Correct pathway choice in relation to the cell cycle is essential to ensure efficient repair, prevention of senescence/apoptosis, and avoiding severe mutations. It is proposed that NHEJ may also function as a failsafe pathway during S-G2 if HR has failed and resected ends are present (Symington LS et al. 2002).

In *Saccharomyces cerevisiae*, NHEJ is initiated when the Ku complex (Yku70-Yku80) binds to the broken DNA ends preventing DNA degradation and 5' resection, which

can lead to HR initiation. Subsequently, Ku complex recruits other NHEJ repair factors including Nej1 and Dnl4/Lif1 which in turn further stabilize Ku on the DNA ends (Chen X and Tomkinson AE 2011 and Yano K et al. 2008). The MRX complex, comprised of Mre11-Rad50-Xrs2, is recruited to break sites and facilitates end-bridging before Dnl4/Lif1 and Nej1 initiate the final ligation step (Yang H et al. 2015). However, the aforementioned events are controlled by complex regulatory networks regulated by DNA damage checkpoints, sensors and chromatin related proteins. This complex networks which coordinate the repair process are highly conserved from single cell organism such as yeast to complex mammalian systems (Finn and Lowndes, 2012).

In *S. cerevisiae*, checkpoints function at the G1 to S transition, intra S-phase, and the G2 to M phase transition. Three kinase-related (PIKKs), primarily Mec1 but also Tel1 function to activate downstream kinases Chk1 and Rad53 as well as promote a general DNA damage response (Stracker TH et al. 2009). This signaling cascade is manifested by a series of phosphorylation events which alter the transcription rate of repair genes and localize proteins responsible for cell cycle progression (Finn K et al. 2012). A review of evidence related to kinases from human cell studies have indicated that a broader phosphorylation landscape involving multiple kinases may facilitate a cross-communication network that spans multiple cellular processes, some of which may not have yet been identified (Bensimon A et al. 2011). Our recent work in the budding yeast model organism suggests that additional kinases may be involved in the broad network of signalling that promotes cell cycle progression following DSBs events (Jessulat M et al. 2015).

In this study we investigated the potential role of the yeast PAS kinase complex (Psk1 and Psk2) in the signalling response to DNA damage. The PAS complex is an evolutionarily

conserved serine/threonine kinase known to play a key role in nutrient partitioning when responding to changing environmental conditions and positive signalling in response to indications of suitable cell integrity (Grose JH et al. 2007). In human cells, PASK has been reported to interact with Rad50, a key protein involved in different DSB repair pathways (Li S. et al. 2014). Here we investigate the role of its yeast homologs, Psk1 and Psk2, in repair of DSBs through NHEJ. The complex has been associated with the cellular response to oxidative stress through the regulation of Sod1. Psk1 and Psk2 are semi-redundant homologs where Psk1 mutants show sensitivity to oxidative stress while Psk2 functions to regulate Ultradian coupled respiratory oscillation (Huang M et al. 2013). Our data suggests that deletion of the PSK1 and PSK2 genes reduce the repair efficiency of DSBs induced on a chromosomal level. We further show that repair of breaks with cohesive and non-cohesive ends are compromised in the absence of PAS kinase complex. Interestingly, *psk1*Δ and *psk2*Δ mutants reduce the accuracy of DSB repair, playing a role in repair fidelity. This previously unreported role may function together or independently of oxidative stress response pathways and vitality signalling pathways.

5.3 Materials and Methods

5.3.1 Strains and plasmids

Deletion strains were in BY4741 background obtained from *S. cerevisiae* deletion library (MATa *orf*Δ::KanMX4 *his3*Δ *leu2*Δ *met15*Δ *ura3*Δ) (Winzler EA et al. 1999). PCR method was used to delete PSK1 and PSK2 in Y7092 (MATa *can1*Δ::STE2pr-HIS3 *lyp1*Δ *ura3*Δ *leu2*Δ *his3*Δ *met15*Δ) and JKM139 (MATa *hmr*Δ::ADE1 *hml*Δ::ADE1 *ade1-100* *leu2-3,112* *lys5* *trp1*::hisG *ura3-52* *ade3*::GAL-HO) background as in (Tong AH and Boone

C 2007). Plasmid p416 contains URA3 marker and Amp^r gene (Jessulat M et al. 2008) and Ycplac111 plasmid contains Leu2 marker and Amp^r (Bahmed K et al. 2010). Plasmids in pGEH background were used in suppression assay, genes are linked to Gal1 promoter with a 2 micron origin of replication as explained (Sopko R et al. 2006).

5.3.2 Plasmid repair assay

Plasmids Ycplac111 and P416 were digested by SmaI and XbaI at their unique site, which lacks homology to chromosomal DNA (Jessulat M et al. 2008 and Yu L and Volker MR 2013). Digested and circular plasmids were transformed into mutants and WT strains in parallel, and grown on selective synthetic media (lacking Ura or Leu). Colony count was performed to obtain number of viable cells for linear and circular plasmid transformations. Ratio of linear to circular transformation was calculated to get the repair efficiency of mutants and it was normalized to the repair efficiency of WT.

5.3.3 Repair fidelity assay

To measure accuracy of repair plasmid based fidelity assay was performed. In this assay, Plasmid pMV1328 was linearized at its KanMX6 marker using PstI. Linearized plasmids were transformed in parallel into mutant and WT strains in JKM139 background. After two to three days of growth on synthetic media without Leu 50-100 colonies were re-grown on YPD media containing G418. Viable cells represented accurate repair, and the ratio of viable to non-viable cells was normalized to the ratio of WT, giving the accuracy of repair of each mutant strain.

5.3.4 Drug sensitivity tests

Mutants and WT strains were grown to saturation in YPD media for 2 day and serially diluted by factor of 10. Each dilution was then spotted on YPD media containing 3 ul/ml phleomycin or 70 mM hydroxyurea (HU) and drug free media as control. To test UV sensitivity cells were spotted on YPD and exposed to UV for 30 seconds. Cell viability of each strain was examined after 2-3 days of growth at 30° C.

5.3.5 Synthetic genetic array analysis

Synthetic genetic array analysis (SGA) was carried out to map the genetic interaction of *PSK1* and *PSK2* as described in (Tong AH and Boone C 2007). Mutant/overexpression strains of *PSK1* and *PSK2* were constructed in α -mating type and crossed to DNA damage array (DDR) and control array. After a series of selections, progenies carrying double mutant or single mutant and overexpression of the candidate genes were obtained. Single mutant with overexpression strains were grown on drug media to examine suppression under DNA damaging conditions. Fitness of double mutant was examined visually, identifying sick and lethal cells. Each experiment was repeated three times.

5.3.6 Chromosomal repair assay

Chromosomal break assay was performed using JKM139 strain. This strain contains a site specific HO-endonuclease linked to GAL-promoter which is activated in the presence of galactose. HO-endonuclease generates DSBs at the MATa locus, though homologous regions to MATa, HML and HMR, are deleted in this strain restricting repair to NHEJ (Jessulat M et al. 2008). WT and mutant strains were grown to saturation and serially diluted by factor of 10. Each dilution was spotted on YPD and YP-Gal media and growth was examined visually after 2 days.

5.4 Results

5.4.1 Plasmid repair analysis suggest a role for *PSK1* and *PSK2* in NHEJ

Plasmid based repair assays have been extensively utilized to investigate novel genes involved in NHEJ repair pathway (Jessulat et al 2008 and 2015, Omid et al 2013). In these assays plasmids are digested with different restriction enzymes producing breaks with different overhangs or blunt ends. This allows studying role of genes in repair of various breaks types through NHEJ. We subjected deletion mutant strains for *PSK1* and *PSK2* to plasmid repair assay, examining the role of *PSK1* and *PSK2* in repair of blunt ended breaks and breaks containing overhangs. Basically, plasmids were digested with restriction enzymes generating breaks with blunt ends or overhangs. Subsequently linearized and intact plasmids were transformed into mutants and WT strains in parallel. Normalized ratio of linearized to circular plasmid of mutant to WT indicated the repair efficiency. Our results showed that *psk1Δ* and *psk2Δ* mutants reduced the repair efficiency of breaks with overhangs to 50% and 47%, respectively compared to WT (Fig. 5.1). This suggests the two PAS kinase proteins may play a role in NHEJ repair. We further investigated their role in repair of blunt ended breaks. Our data shows that *psk1Δ* and *psk2Δ* mutant strains were also deficient in repair of breaks with blunt end, reducing repair efficiency to 49.5% and 41% respectively (Fig 5.1). Based on these data, we suggest that *PSK1* and *PSK2* are implicated in repair of different types of break through NHEJ.

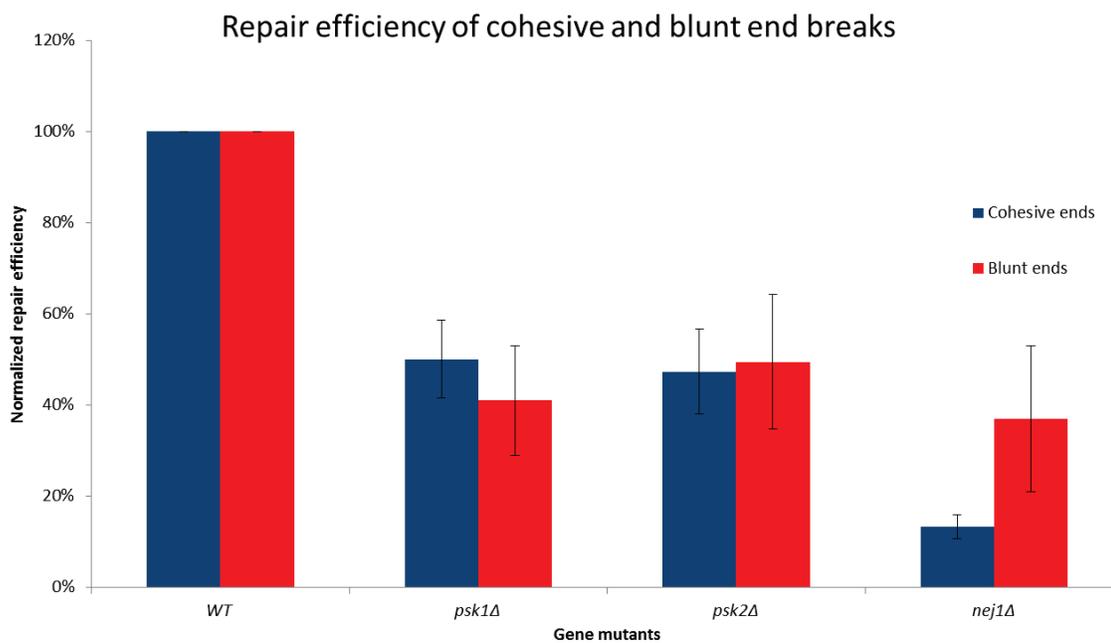


Figure 5.1: Deletion of PSK1 and PSK2 reduced repair efficiency of breaks with cohesive and non-cohesive ends in a plasmid based assay. Experiments were repeated 5 times.

5.4.2 Deletion of *PSK1* or *PSK2* affects the fidelity of DSB repair

NHEJ pathway is associated with mutagenic repair, particularly in case of lesions requiring end processing and resection due to presence of damaged nucleotides (Bétermier M et al. 2014). Several studies have identified genes which alter the repair fidelity through NHEJ (Bahmed K et al. 2010, Bahmed K et al. 2011 and Yu L and Volkert ML 2013). Therefore we investigated the role of *PSK1* and *PSK2* in carrying out accurate repair via a plasmid based assay. In brief, plasmids are digested at KanMX6 gene and transformed into mutant and WT strains. In the first step, cells are grown on selective media to screen for successful repair of plasmids. Viable cells are then grown on media containing G418 to score for repair accuracy. Our data showed that the accuracy of repair in cells carrying *psk1Δ* is 55% lower than WT (fig 5.2). Furthermore, 59% of the repaired plasmids were mutagenic in *psk2Δ* mutant strains compared to WT (Fig. 5.2). These data suggest that *PSK1* and *PSK2* are implicated in promoting an error-free repair through NHEJ.

5.4.3 Chromosomal break analysis confirms the influence of *PSK1* and *PSK2* on the efficiency of NHEJ

Since the plasmid based assay showed repair deficiency for *psk1Δ* and *psk2Δ* mutant strains, we investigated their role in repair of chromosomally induced breaks. In JKM139 strain, a GAL-promoter linked HO-endonuclease is activated in the presence of galactose, generating DSBs at the MATa locus. In order to restrict the repair to NHEJ, HML and HMR, which are the homologous regions of MATa are deleted in this strain (Jessulat M et al 2007). Deletion strains of *PSK1* and *PSK2* as well as WT were grown to saturation and spotted on

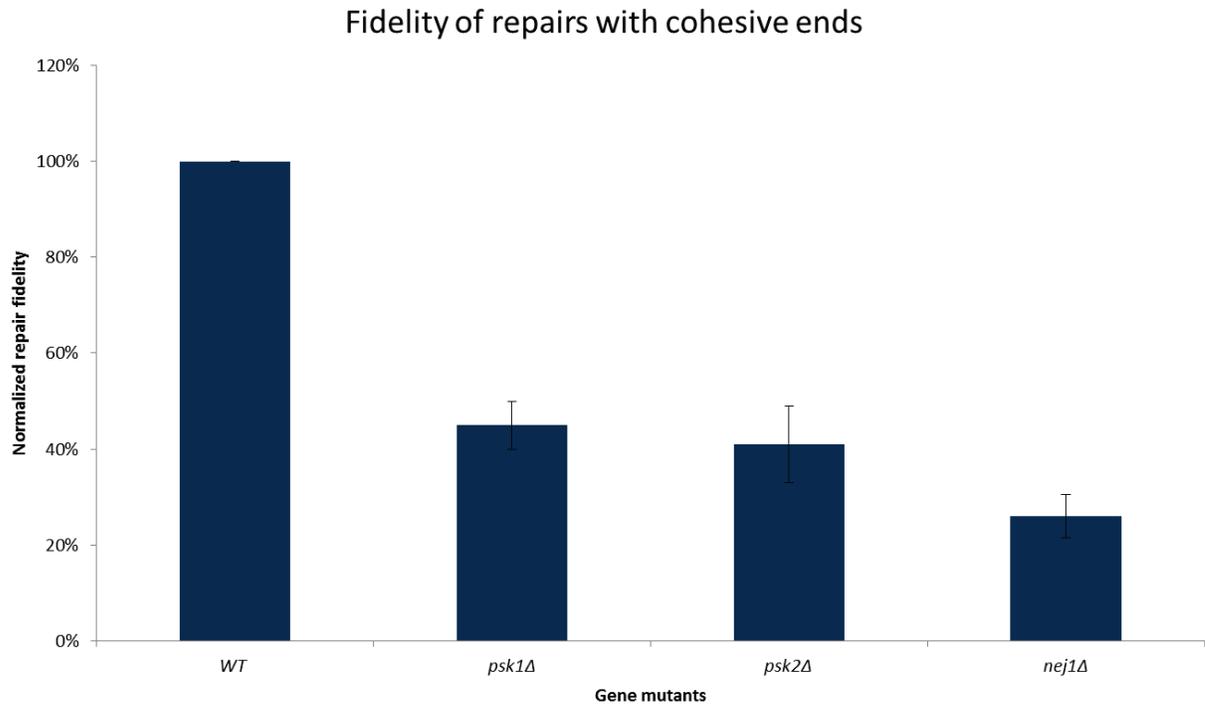


Figure 5.2: Psk1 and Psk2 are important in carrying out non-mutagenic repair. Deletion of both genes reduces repair accuracy compared to WT.

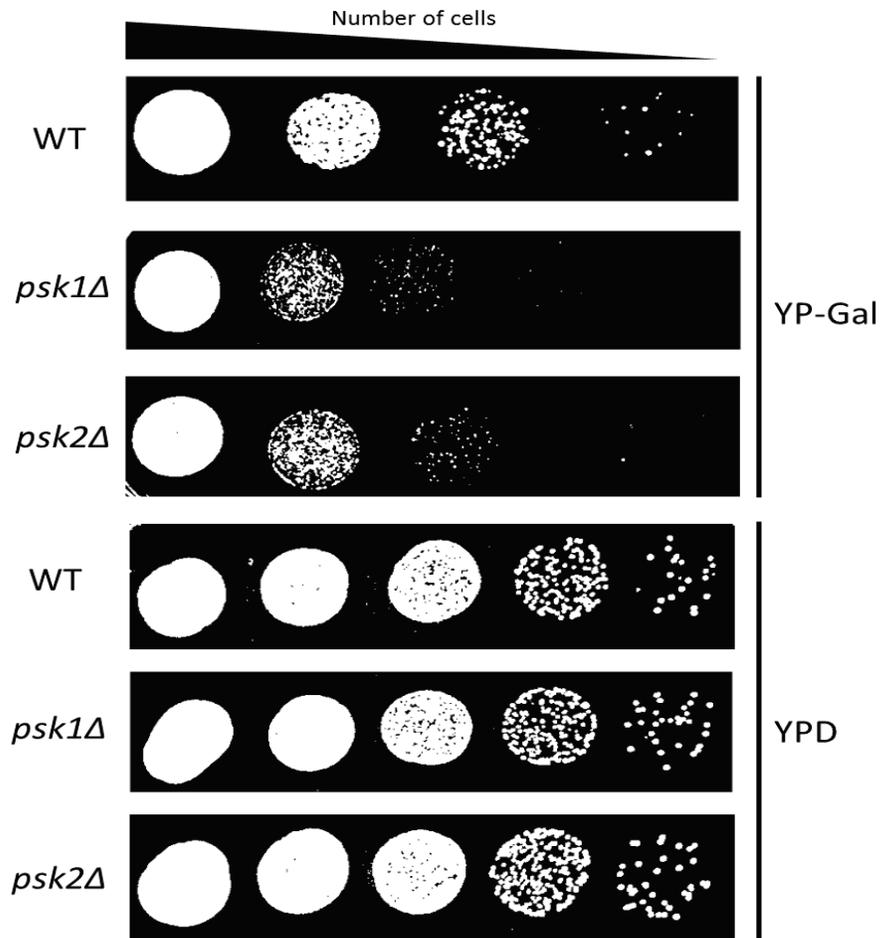


Figure 5.3: Chromosomal breaks are induced by HO-endonuclease in JKM139 strains. Deletion of *PSK1* and *PSK2* reduced viability under DSBs induction.

media containing glucose and galactose. Cell viability of mutant strains was used as a measure of repair efficiency in comparison to WT strain. Based on our data, *psk1Δ* and *psk2Δ* mutants reduce the efficiency of repair compared to WT (Fig. 5.3). This data is in accordance with the plasmid based repair assay further supporting a role for *PSK1* and *PSK2* in repair of DSBs via NHEJ.

5.4.5 Drug sensitivity profiles links *PSK1* and *PSK2* to DNA repair

Cell sensitivity to DNA damaging agents is an indication of defects in DNA damage response (DDR) or DNA repair pathway. Deletion of genes involved in DDR or repair increases sensitivity to DNA damaging agents therefore drug sensitivity has been used to screen for genes involved in different repair pathways (Birrell GW et al. 2001). Generally DNA damaging agents induce various types of DNA lesions at random sites on the chromosome, hence they often don't target a specific repair pathway. In this study to examine sensitivity of *psk1Δ* and *psk2Δ* to DNA damaging agents we used HU and PHLM. HU induces DNA breaks by nucleotide depletion which results in collapse of replication machinery (Koç A et al. 2004). PHLM promotes production of free radicals inducing different types of DNA lesions, including DNA single strand breaks (SSBs) and DSBs, while UV radiation promotes formation of pyrimidine dimers (Birrell GW et al. 2001 and Povirk LF et al. 1981). Mutants and WT strains were grown to saturation and spotted on media containing drug and drug free media. To test UV sensitivity cells were spotted on drug free media and exposed to UV light for 30 seconds. Our results showed that *psk1Δ* mutants are only sensitive to UV radiation, while *psk2Δ* mutants are sensitive to HU and PHLM (Fig. 5.4). These observations further support involvement of *PSK2* in DNA damage repair.

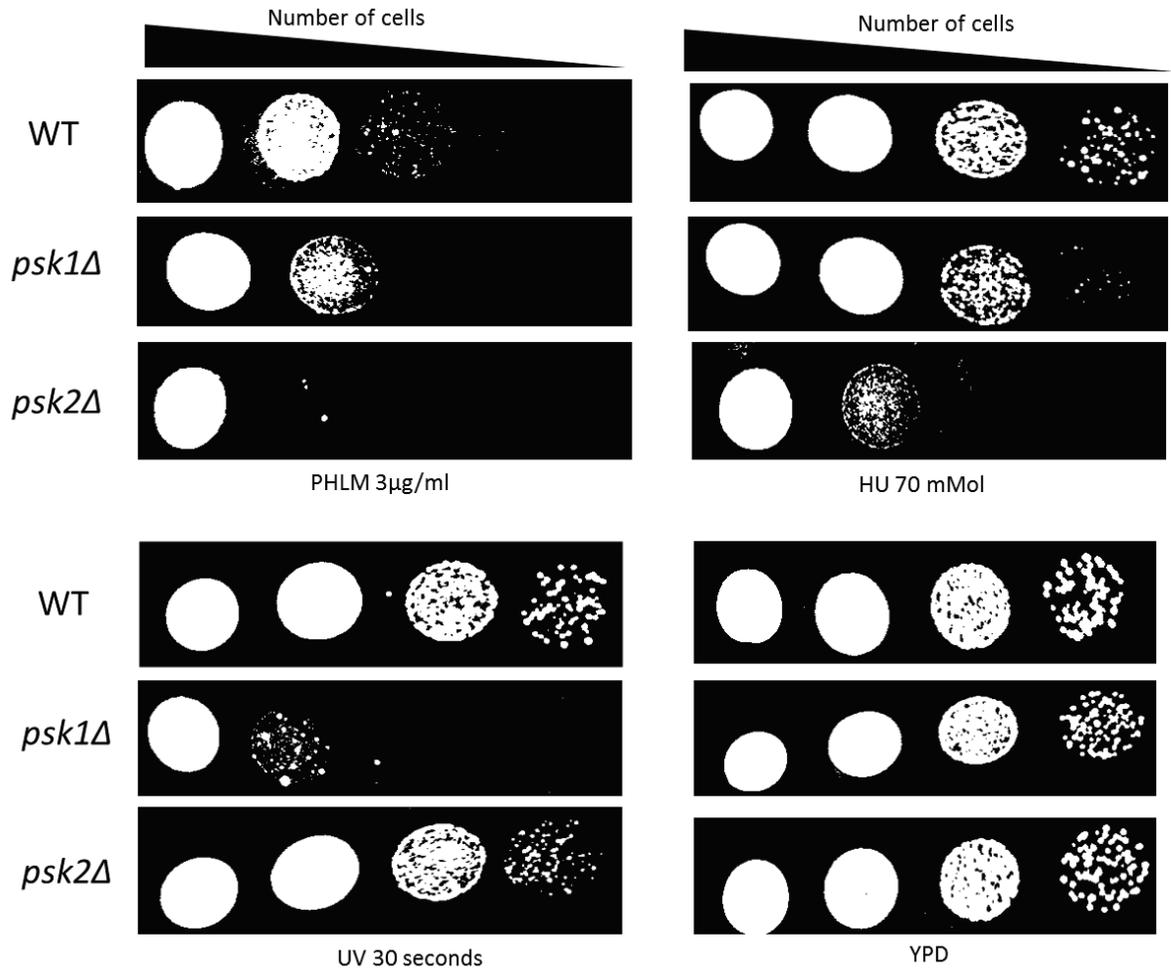


Figure 5.4: Deletion of *PSK1* and *PSK2* showed different profile in sensitivity to DNA damaging agents. Mutant strains of *PSK1* were only sensitive to UV radiation, whereas *psk2Δ* mutants showed sensitivity to HU and PHLM.

Since deletion of *PSK2* shows sensitivity to two different DNA damaging agents, it could potentially play a more significant role in the process of DNA repair compared to *PSK1*.

5.4.6 Genetic interaction analysis support a role for *PSK1* and *PSK2* in DNA repair

Genetic interactions (GIs) analysis is a useful mean in understanding the functional relations between different genes. Two genes are considered genetically interacting when deletion of both genes results in a phenotype that significantly differs from deletion of each gene individually. Therefore, deletion of two genes resulting in lower fitness or cell lethality is defined as negative genetic interaction. This form of interaction generally implicates the interacting genes function in parallel pathways (Boone C et al. 2007). To examine the role of *PSK1* and *PSK2*, we carried out miniaturized SGA analysis for the target genes. In this assay, we crossed the query gene mutants in α -mating type to DNA Damage Array (DDA) containing mutant strains for genes functioning in DNA repair, DDR, cell cycle and related processes. As a control, we also crossed our query strains with an array of gene mutants with function in pathways unrelated to DNA damage repair. The resulting progeny carrying double mutants were scored for fitness following a series of selections. GI network for *PSK2* included genes involved in different repair pathways and DNA damage checkpoint proteins. *PSK2* genetically interacts with all the members of MRX complex, involved in HR and NHEJ, several genes involved in HR including *RAD52* and *RAD51*, and mismatch repair (MMR) protein *MSH6* (Fig. 5.5A). Furthermore *PSK2* GI analysis showed interactions with *SWI4* and *SWI6* which function upstream of both repair pathways. *PSK1* GI network included *MRE11*, *LIF1* and *NEJ1* which are key members of NHEJ pathway, *RAD52* which is part of HR, *MSH5* and *MLH1* of mismatch repair and *SWI6* which is involved in DDR (Fig. 5.5A). GI network of *PSK1* and *PSK2* show similar trend in the targeted pathways, both

genes interact with genes involved in HR and MMR. Furthermore, both genes interact with key member of NHEJ, HR and DDR pathway, *MRE11*, as well as HR gene *RAD52*, and *SWI6* a transcriptional regulator activated in response to DNA damage. This is a similar profile to those for key NHEJ genes including *YKU70* and *YKU80* (Koh JLY et al. 2012) and supports a role for *PSK1* and *PSK2* genes in NHEJ. Also, in comparison to *PSK1*, the higher number of interactions observed for *PSK2*, present the possibility of *PSK2* playing a more dominant role in this pathway; this is in agreement with our drug sensitivity data (please see above).

Phenotypic suppression analysis (PSA) investigates genetic interactions by studying effects of overexpression of a gene on deletion of a different gene under stress conditions. These studies reveal the order by which the genes function in a specific pathway. PSA examines the ability of a gene overexpression in rescuing the phenotype of a different gene deletion under specific conditions. These forms of interactions, which result in phenotypic enhancements, are defined as positive interactions. To investigate the effects of *PSK1* and *PSK2* in suppressing deletion of DNA damage related genes, we introduced expression vectors of *PSK1* and *PSK2* into DDA. Subsequently, strains containing expression plasmid in the mutant background were grown in the presence of drugs and their fitness was observed. Overexpression of *PSK1* compensated for deletion of genes involved in NHEJ (*MRE11* and *RTT109*), DNA DDR (*MRC1*, *SWI6*, *ARP8*) and spindle checkpoint (*SGO1*) (Fig 5.5B). Similar trend of interactions was observed for *PSK2* as it suppressed phenotypes of mutant of genes functioning in NHEJ (*MRE11*, *RAD50* and *RTT109*), DDR (*MRC1*, *SWI6* and *DUN1*) and spindle checkpoint (*SGO1*) (Fig 5.5B). PSA network of *PSK1* and *PSK2* have several genes involved in DDR process in common, namely: *MRC1*, *MRE11*, *SWI6*, *RTT109* and

SGOI. These data is in accordance with the GIs obtained from SGA analysis, further linking the roles of Psk1 and Psk2 to DNA damage, potentially upstream of both repair pathways.

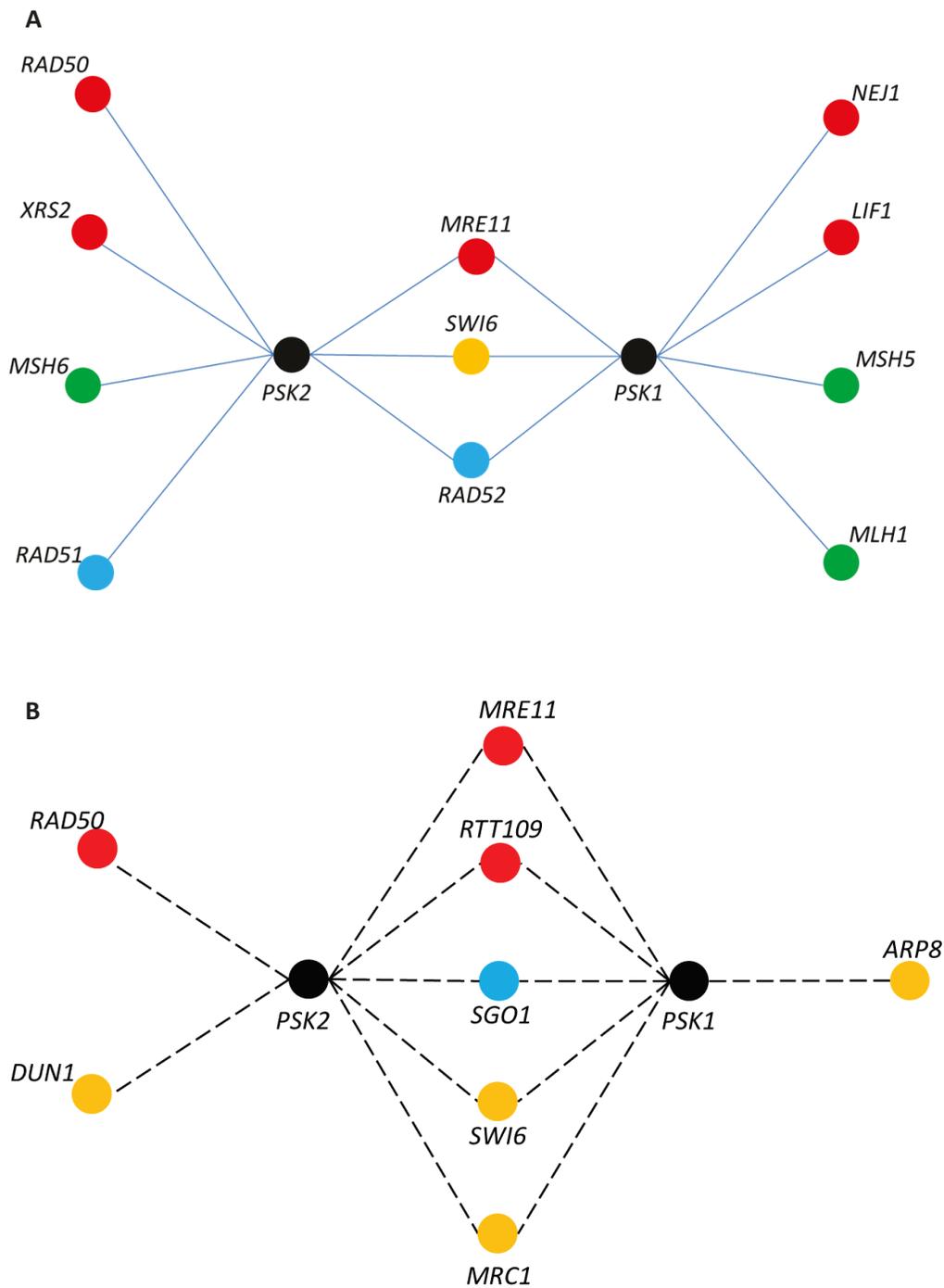


Figure 5.5: Genetic interaction maps of PSK1 and PSK2 showed similar trends with several interacting genes in common A) SGA analysis revealed that PSK1 and PSK2 interact with genes involved in NHEJ (red), HR (blue) and DDR (orange). B) PSA analysis showed under DNA damaging conditions, PSK1 and PSK2 overexpression suppress phenotypes of five common genes, MRC1 and SWI6 of DDR, MRE11 and RTT109 of NHEJ and SGO1 of HR.

5.5 Concluding remarks and future directions

PAS kinase is an evolutionary conserved protein present in human and yeast, with the exception of yeast containing two orthologs (DeMille D and Grose JH 2013). In *S. cerevisiae* the PAS complex is composed of Psk1 and Psk2 and controls metabolic rate and energy homeostasis. The PAS kinase has a protein kinase domain and a PAS domain, suggesting a role in sensory-coupled signal transduction pathways (Grose J et al. 2007). Psk1 and Psk2 carry out different functions in sugar metabolism, suggesting that these orthologs may also work independent of each other (DeMille D and Grose JH 2013). Kinases are important proteins in signaling pathways and play key roles in communications between pathways and within different pathways. Here we report a novel function for Psk1 and Psk2 in DNA repair through NHEJ. Our analysis showed that repair efficiency of breaks with cohesive and non-cohesive ends as well as chromosomally induced breaks are lowered in cells lacking Psk1 and Psk2. Furthermore, deletion of *PSK1* and *PSK2* reduce repair accuracy, supporting a role for both genes in repair fidelity. There is no reported role for Psk1 and Psk2 in DNA damage context, highlighting the novelty of this work.

Future investigation of the relationship between *PSK1*, *PSK2* and DNA damage checkpoint genes may aid in understanding the role of these kinases in DNA repair. GI network of *PSK1* and *PSK2* included genes involved in DDR such as *MRC1*, *MRE11* and *SWI6*. Interestingly both genes when overexpressed compensate for deletion of *MRE11* and *SWI6* and show negative genetic interaction with the same genes. This implicates a role for *PSK1* and *PSK2* related to *MRE11*. Genetic relation of these genes can be investigated by observing the repair efficiency of single mutants (*psk1Δ*, *psk2Δ* and *mre11Δ*) compared to double mutants strains (*psk1Δmre11Δ* and *psk2Δmre11*) in a plasmid based assay.

Furthermore, Kinases have been shown to play a central role in signaling and communication in the DDR and DNA repair pathways. DNA checkpoint proteins may act as a substrates of PAS kinase, therefore it will be of interest to carry out phosphorylation assay to investigate the role of Psk1 and Psk2 in targeting the candidate genes obtained from the genetic analysis.

Chapter 6: Conclusion and future directions

6.1 Final thoughts and concluding remarks

The pathways leading to DNA double stranded breaks (DSBs) repair are highly conserved processes essential in maintaining genomic integrity and identity. Defects in the process of DSBs repair endanger genomic integrity by introducing various alterations to the chromosomes. In addition to local mutations, DSBs contribute to genomic instability through chromosomal rearrangement and gene translocation which are characteristics of cancer cells (Aparicio T et al. 2014). Genomic instability is a common characteristic of cancer cells. It assists in the onset of cancer and inhibition of apoptosis in tumor cells (Shibata A and Jeggo PA 2014). Functional DNA damage response (DDR) pathway is crucial in monitoring chromosomal integrity and inhibition of carcinogenesis through detection of DNA damage and initiation of DNA repair. One of the common characteristics of cancer cells is impairment of DNA repair pathways; particularly repair of highly genotoxic DSBs (Kelley MR et al. 2014). In fact, mutations in genes involved in key repair pathways of DSBs, homologous recombination (HR) and non-homologous end joining (NHEJ) have been implicated in the onset of cancer (Shrivastav M et al. 2008). Mutations in ATM, CHK2, RAD51, BRCA1 and BRCA2, part of DDR and HR pathways, increase susceptibility to breast, ovarian and lung cancer (reviewed in Dietlein F et al. 2014). Furthermore, mutations in NHEJ proteins DNA PK and 53BP1 are linked to development of pancreatic and breast cancer (Aparicio T et al. 2014, Wiegmans AP et al. 2014 and Hsu F-M et al. 2012). Several studies have shown that cells lacking the activity of key NHEJ proteins, Ku70, Ku80 and DNA PKcs, show increased levels of chromosome translocation which is a key factor in the onset of cancer (Davis AJ and Chen DJ 2013 and Bunting SF and Nussenzweig A 2013).

In addition to its role in maintaining genomic stability, NHEJ is essential to development of immune response via Variety, Diversity and Joining, or V(D)J pathway. In this process, endonucleases induce DSBs nearby the conserved region of antigen receptor genes and between the variant genes V (Variety), D (Diversity) and J (Joining). End-joining between the conserved regions and the variant regions is carried out by NHEJ machinery producing different types of antigen receptors (Deriano L and Roth DB 2013). Efficient V(D)J recombination requires key members of NHEJ pathway including Artemis, DNA PK, XLF, Ku70, Ku80, Lig4 and Xrcc4 (Roy S et al. 2012, Zha S et al. 2012 and Liu X et al. 2012). Mutations in DNA PK, Artemis, Lig4 and XLF have been observed in patients with immunodeficiency syndromes, linking NHEJ with immune response disorder (reviewed in Woodbine L et al 2014). Furthermore, dysfunctional genes involved in NHEJ process have been associated with neurodegenerative diseases (Rass U et al. 2007). Association of NHEJ and DNA repair with cancer development, immune response and neuron disorders further highlights the importance of understanding the repair mechanism pathway.

The network of cellular processes involved in the response to DSBs highlights complexity of the repair process in the cell. Response to DSBs requires alterations to the chromatin structure by chromatin remodeling complexes and histone modifiers and changes in gene expression. It also results in stop or delay in the cell cycle progression requiring number of signaling kinases and adaptor proteins (Attikum H.V and Gasser S.M 2009). This signifies the number of factors involved in the cross-communication between NHEJ and other cellular processes. Identification of novel players involved in the repair pathway facilitates better understanding of mechanisms and regulations that underlie this process.

DSBs repair and NHEJ pathway are highly conserved from *S. cerevisiae* to human, making yeast an ideal model organism to investigate DNA repair. The key proteins involved in the process, KU heterodimer, Lig4/Lif1 (Xrcc4), and Nej1 (XLF) have been identified in human and yeast and have been the subject of vigorous investigations for a number of years. Though in recent years additional genes have been linked to the repair pathway. These genes are shown to play various roles in the process of NHEJ for example: *PPH3/PSY2* connects NHEJ to DNA damage checkpoints (Omidi et al. 2014), *BUB1/BUB2* regulates cell cycle progression during NHEJ repair (Jessulat M et al. 2015) and *TBF1/VID22* alters chromatin structure at DSBs and promotes end resection in NHEJ (Bonetti D et al. 2013). Identification of additional NHEJ factors indicates the possibility of other novel genes that can influence the efficacy of NHEJ. It illustrates that the process as well as its control and regulations is not fully understood and there is a need to further investigate novel genes that affect this process. These genes may play regulatory roles within NHEJ or be involved in the networking and communication between NHEJ and other processes.

In recent year systematic approaches adopted by functional genomics studies have been employed to identify novel genes in different process (Alamgir et al. 2008, Samanfar B et al. 2013 and 2014 and Vlasblom J et al. 2015). In the context of DDR and DNA repair, several studies have been carried out on a genome-wide scale to identify novel genes, by studying sensitivity to DNA damaging agents, protein-protein interactions (PPI), gene expression profiles and genetic interactions (GI) (Brown JA et al. 2006, Chang M et al. 2002, Hendry JA et al. 2015, Bandyopadhyay S et al. 2010). Besides experimental methods, computational tools have been utilized in functional genomics studies to identify genes with novel functions (Schoenrock A et al. 2014). Bioinformatics approaches can carry out

genome-wide predictions in a timely manner with low cost. A drawback of bioinformatics tools is that even though data obtained by such prediction tools are reported to have up to 70% success rate, they always require further confirmation through traditional experimental methods to validate their accuracy (Pitre S et al. 2012, Jessulat M et al. 2011). In computational methods, precision of the generated data is reflected by sensitivity and specificity which measure the rate of true positives and true negatives, respectively. Such global studies often generate large amount of data highlighting the importance of obtaining high specificity to eliminate false positives. Sensitivity and specificity are often reversely proportional and increasing specificity generally results in lower sensitivity, reducing the coverage. In our study, total possible interactions were 650,000; therefore it was important to carry out the experiment at a very low rate of false positives such as 99.95%. At this low rate the number of false positives is estimated to be 325 which seem minimal compared to the total number of possible interactions. Novel interactions obtained by our study added a notable number of candidates to the growing list of proteins with potential role in DNA repair pathway in human cells. Experimental validation of our results showed 50% success rate of predictions through an indirect functional assay, proving PIPE to be a powerful tool in screening for novel protein function. These novel candidates could be further investigated to identify their role in DNA repair, adding substantial information to our knowledge of the repair pathway and protein functional network of the cell.

We identified seven novel genes which reduce the efficiency of repair through our plasmid based repair assay, four of which were subjected to further experimentation to study their role in NHEJ. Aside from reducing the efficiency of repair, we showed three of these genes regulate repair fidelity in NHEJ. Fidelity of repair is essential to the cell, as inaccurate

repair can lead to mutagenesis promoting genomic instability and causing complications for the cell. Despite it being considered an “error-prone” pathway, recent studies have shown that fully functional NHEJ process generally carries out the repair process with high accuracy (Bétermier M et al. 2014). It is proposed that generation of errors during repair is attributed to alt-NHEJ which is activated in the absence of key NHEJ factors (Pannunzio NR et al. 2014 and Deriano L and Roth DB 2013). The alt-NHEJ pathway uses micro-homology for repair, though it results in deletions within the repaired segment compromising genetic integrity. Identification of genes involved in repair fidelity can help in elucidating the poorly understood alt-NHEJ pathway.

In this thesis, we followed up our investigations with four genes which have different reported functions namely: *TPK1* a kinase functioning in stress response, *ARP6* playing a role in gene expression and *PSK1/PSK2* kinases involved in sugar metabolism. It must be noted that none of the candidate genes identified in our study had been previously linked to NHEJ or DSBs repair, highlighting the novelty of this work. We showed these genes are involved in different aspects of the repair process. They are either involved directly through affecting key NHEJ factors, as was the case for *TPK1*, or hypothesized to regulate the process, thought to be the case for chromatin remodeling factor *ARP6*. Role of Tpk1 in NHEJ is connected to Nej1, a key protein involved in the final ligation step. This is an important finding and suggests the presence of a gap of knowledge in understanding the mechanism of NHEJ. Recruitment of proteins to the site of damage and activity of the key factors of NHEJ depends on various elements. Major NHEJ proteins regulate recruitment of others, for example recruitment of Lif1 depends on Yku80 and Xrs2 (Wu D et al. 2008); therefore identification of novel proteins, such as Tpk1 functioning in recruitment of a key NHEJ also

aids in elucidating the communication within the pathway. Furthermore, chromatin remodeling complexes INO80, SWR1 and RSC as well as checkpoint proteins Tel1, Mec1 and Rad53 play crucial roles in recruitment and regulation of key members of the repair pathway (Bao Y and Shen X 2007, Ataian Y and Kerbs EJ 2006 and Finn K et al.2012). We suggest that *ARP6* regulates NHEJ through its association with *RSC2*, a chromatin remodeler with various roles in DNA repair and DDR. Altogether our observations suggest there are many aspects of the DNA repair process that must be studied to better understand the functional network underlying DDR process.

6.1.1 Future prospect

Human homolog of *TPK1*, *PRKCAB*, was shown to reduce repair efficiency, further indicating the conservation in NHEJ pathway. Furthermore, *ARP6* and *PSK* proteins have human homologs which may potentially be involved in the repair pathway as both homologs, *ACTL6B* and *PASK*, also interact with repair proteins. NHEJ is an important factor in immunodeficiency syndrome and cancer development, therefore these novel repair proteins can be used as potential drug targets. A driving force in cancer cell survival is defects in DNA repair pathways which promote genomic instability. Defective repair pathway in cancer cells forces the cell to rely on alternative secondary pathways for repair, which are generally mutagenic (Pearl LH et al. 2015). Although impairment of DNA repair pathway gives cancer cells a competitive advantage by inhibiting apoptosis, cell's weakness in response to DNA damage has been exploited by researchers in designing therapeutic drugs (Pearl LH et al. 2015). Several drugs used in chemotherapy and radiation therapy target DDR pathway either by generating DNA lesions directly or blocking replication forks causing DNA damage (Powell SN and Bindra RS 2009). Examples of these drugs include: bleomycin which breaks

DNA backbone, cisplatin which crosslinks DNA strands and temozolamide which is an alkylating agent (Pearl LH et al. 2015 and Della Latta V et al. 2015 and Dietlein F et al 2014). These drugs target the general pathway of DDR, though recently drugs have been designed which are specific to a protein in the DDR or DNA repair pathway. Currently there are several drugs in clinical development, such as olaparip which inhibits Poly (ADP-ribose) Polymerase (PARP) involved in base excision repair (BER), CC-115 which inhibits DNA PK functioning in NHEJ, and AZD0156 which is an inhibitor of ATM a key DDR signaling kinase (Powell SN and Brinda RS 2009 and Dietlein F et al. 2014). In case of breast and ovarian cancer, in which HR is compromised, DNA PK inhibitors are shown to have promising therapeutic application (Dungl DA et al. 2015). Another NHEJ protein with promising therapeutic use is DNA Ligase IV, inhibition of which is cytotoxic to radiation in mouse model (Srivastava M et al. 2012). These targeted inhibitors can also be used in combination with traditional chemotherapy to enhance the effects of the therapy. These findings indicate the importance of identifying novel protein which could be used as targets for therapeutic purposes. In theory, the novel proteins identified in the current thesis could be used as possible targets for the development of new drugs.

In this study we employed PPIs prediction to identify potential candidates playing a role in NHEJ pathway. Genetic interactions (GIs) analysis can also be utilized to generate a list of novel candidates involved in the repair pathway. The comprehensive map of yeast GI network has been reported using synthetic genetic array analysis and its publicly available (Tong AH et al. 2006 and Koh JLY et al. 2010). Other studies have targeted specific pathway, identifying the genetic interactions within DDR or repair pathways, providing insights into the dynamic of the DDR pathway (Bandyopadhyay S et al 2010). Useful

information could be obtained from these large scale analyses to identify novel gene candidates involved in different repair pathways.

The computational tool that we utilized in this thesis uses the primary sequence of polypeptide for PPIs prediction. This method can be improved to apply other properties of proteins, such as structure and expression, to enhance the accuracy and precision of the predicted PPIs. In the current study we have only looked at a subset of PPIs in human cell, this data could be expanded to obtain a more comprehensive coverage of the human proteome. In addition, it will be interesting to experimentally examine the predicted interactions validating the PPIs by tandem affinity purification under native and DNA damage conditions.

Novel candidates reported in this study ARP6 and PSK1/PSK2 have human homologs which interact with known DNA repair genes. A follow up experiment can be conducted to investigate the role of these homologs in NHEJ in mammalian cell lines. These studies can further examine conservation in the function of the novel candidates in human compared to yeast. Role of PSK1/PSK2 in NHEJ is not fully understood. Additional experiments can be designed to elucidate the role of PSK1/PSK2 in the repair process. Future work can include: directed kinase assay to find potential DNA damage targets, genetic analysis of PSK1/PSK2 with other DNA damage checkpoints to investigate functional redundancy between the genes and chromatin immunoprecipitation to investigate recruitment to the site of damage. Furthermore, human homolog of Psk proteins, PASK, interacts with Rad50. To examine potential conservation of this interaction, yeast 2 hybrid assay can be performed.

Identification of new genes provides substantial information not only in the process of DNA repair, but also in understanding of the cell as a system. Reported genes have different known functions, which brings up the possibility of them being involved in the cross-communication between NHEJ and other cellular pathways. Understanding of such cross-communications requires detailed investigation of the protein activity in response to DNA damage.

References

- Åberg E, Lund B, Pflug A, et al. Structural origins of AGC protein kinase inhibitor selectivities: PKA as a drug discovery tool. *Biological Chemistry*. 2012;393(10):1121-9.
- Aggarwal K, Lee KH. Functional genomics and proteomics as a foundation for systems biology. *Brief Functional Genomic Proteomic*. 2003;2(3):175-84.
- Ahnesorg P, Jackson SP. The non-homologous end-joining protein Nej1p is a target of the DNA damage checkpoint. *DNA repair*. 2007;6(2):190-201.
- 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.
- Alamgir M, Erukova V, Jessulat M, Azizi A, Golshani A. Chemical-genetic profile analysis of five inhibitory compounds in yeast. *BMC Chemical Biology*. 2010;10:6.
- Aloy P, Russell RB. Structural systems biology: modelling protein interactions. *Nature Review Molecular Cell Biology*. 2006;7(3):188-97.
- Amos-Binks A, Patulea C, Pitre S, et al. Binding Site Prediction for Protein-Protein Interactions and Novel Motif Discovery using Re-occurring Polypeptide Sequences. *BMC Bioinformatics*. 2011;12:225.
- Andres SN, Vergnes A, Ristic D, Wyman C, Modesti M, Junop M. A human XRCC4–XLF complex bridges DNA. *Nucleic Acids Research*. 2012;40(4):1868-1878.
- Anghileri P, Branduardi P, Sternieri F, et al. Chromosome separation and exit from mitosis in budding yeast: dependence on growth revealed by cAMP-mediated inhibition. *Exp Cell Res*. 1999;250(2):510-23.
- Aparicio T, Baer R, Gautier J. DNA double-strand break repair pathway choice and cancer. *DNA repair*. 2014;19:169-175.
- Ataian Y, Kerbs EJ. Five repair pathways in one context : chromatin modification during DNA repair. *Biochem Cell Biology*. 2006;84:490-504.
- Attikum HV, Fristch O, and Gasser SM. Distinct role for Swr1 and Ino80 chromatin remodeling complexes at chromosomal double-strand break. *EMBO journal*. 2007a;26:4113-4125.
- Attikum HV, Fritsch O, Gasser SM. Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks. *The EMBO Journal*. 2007b;26(18):4113-4125.
- Attikum HV, Gasser SM. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol*. 2009;19(5):207-17.

- Babu M, Arnold R, Bundalovic-Torma C, et al. Quantitative Genome-Wide Genetic Interaction Screens Reveal Global Epistatic Relationships of Protein Complexes in *Escherichia coli*. Sasseti CM, ed. *PLoS Genetics*. 2014;10(2):e1004120.
- Babu M, Díaz-Mejía JJ, Vlasblom J, et al. Genetic Interaction Maps in *Escherichia coli* Reveal Functional Crosstalk among Cell Envelope Biogenesis Pathways. Burkholder WF, ed. *PLoS Genetics*. 2011;7(11):e1002377
- Bahmed K, Nitiss KC, Nitiss JL. Yeast Tdp1 regulates the fidelity of nonhomologous end joining. *PNAS*. 2010;107(9):4057-4062.
- Bahmed K, Seth A, Nitiss KC, Nitiss JL. End-processing during non-homologous end-joining: a role for exonuclease 1. *Nucleic Acids Research*. 2011;39(3):970-978.
- Bandyopadhyay S, Mehta M, Kuo D, et al. Rewiring of Genetic Networks in Response to DNA Damage. *Science (New York, NY)*. 2010;330(6009):1385-1389.
- Bao Y, Shen X. Chromatin remodeling in DNA double-stranded break repair. *Current Opinion in Genetics and Development*. 2007;17:126-131.
- Bellay J, Atluri G, Sing TL, et al. Putting genetic interactions in context through a global modular decomposition. *Genome Research*. 2011;21(8):1375-1387.
- Ben-Hur A, Noble WS. Choosing negative examples for the prediction of protein-protein interactions. *BMC Bioinformatics*. 2006;7(Suppl 1):S2.
- Bensimon A, Aebersold R, Shiloh Y. Beyond ATM: the protein kinase landscape of the DNA damage response. *FEBS Letters*. 2011;585(11):1625-39.
- Berggård T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. *Proteomics*. 2007;7(16):2833-42.
- Bétermier M, Bertrand P, Lopez BS. Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? Jinks-Robertson S, ed. *PLoS Genetics*. 2014;10(1):e1004086.
- Birrell GW, Giaever G, Chu AM, Davis RW, Brown JM. A genome-wide screen in *Saccharomyces cerevisiae* for genes affecting UV radiation sensitivity. *PNAS*. 2001;98(22):12608-12613.
- Boiteux S, Guillet M. Abasic sites in DNA: repair and biological consequences in *saccharomyces cerevisiae*. *DNA Repair (Amst)*. 2004;3(1):1-12.
- Bonetti D, Anbalagan S, Lucchini G, Clerici M, Longhese MP. Tbf1 and Vid22 promote resection and non-homologous end joining of DNA double-strand break ends. *EMBO*. 2013;32(2):275-289.
- Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nat Review Genetics*. 2007;8(6):437-49.

- Boulton SJ, Jackson SP. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Research*. 1996;24(23):4639-4648.
- Branzei D and Foiani M. The Rad53 signal transduction pathway: Replication fork stabilization, DNA repair, and adaptation. *Experimental Cell Research*. 2006;312:2654–2659.
- Brown JA, Sherlock G, Myers CL, et al. Global analysis of gene function in yeast by quantitative phenotypic profiling. *Molecular Systems Biology*. 2006;2:2006.0001.
- Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *International Journal of Molecular Sciences*. 2009;10(6):2763-2788.
- Bunting SF, Nussenzweig A. End-joining, translocations and cancer. *Nature Review Cancer*. 2013;13(7):443-54.
- Butland G, Peregrín-Alvarez JM, Li J, et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*. 2005;433(7025):531-7.
- Cagney G, Alvaro D, Reid RJ, Thorpe PH, Rothstein R, Krogan NJ. Functional genomics of the yeast DNA-damage response. *Genome Biology*. 2006;7(9):233.
- Callebaut I, Malivert L, Fischer A, Mornon JP, Revy P, de Villartay JP. Cernunnos interacts with the XRCC4 x DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor Nej1. *Journal Biological Chemistry*. 2006;281(20):13857-60.
- Chambers AL, Brownlee PM, Durley SC, Beacham T, Kent NA, Downs JA. The Two Different Isoforms of the RSC Chromatin Remodeling Complex Play Distinct Roles in DNA Damage Responses. Bielinsky A-K, ed. *PLoS ONE*. 2012;7(2):e32016.
- Chang M, Bellaoui M, Boone C, Brown GW. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *PNAS*. 2002;99(26):16934-16939.
- Chang M, Bellaoui M, Boone C, Brown GW. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *PNAS*. 2002;99(26):16934-16939.
- Chapman JR, Taylor MRG, Boulton SJ. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell*. 2012 47:497-510
- Chen L, Trujillo K, Ramos W, Sung P, Tomkinson AE. Promotion of Dnl4-Catalyzed DNA End-Joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 Complexes. *Molecular Cell*. 2001;8:51105-1115.

- Chen X, Tomkinson AE. Yeast Nej1 Is a Key Participant in the Initial End Binding and Final Ligation Steps of Nonhomologous End Joining. *The Journal of Biological Chemistry*. 2011;286(6):4931-4940.
- Chia J-M, Kolatkar PR. Implications for domain fusion protein-protein interactions based on structural information. *BMC Bioinformatics*. 2004;5:161.
- Chien CT, Bartel PL, Sternglanz R, Fields S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *PNAS*. 1991;88(21):9578-9582.
- Chiruvella KK, Liang Z, Wilson TE. Repair of Double-Strand Breaks by End Joining. *Cold Spring Harbor Perspectives in Biology*. 2013;5(5):a012757.
- Chiruvella KK, Renard BM, Birkeland SR, Sunder S, Liang Z, Wilson TE. Yeast DNA ligase IV mutations reveal a nonhomologous end joining function of BRCT1 distinct from XRCC4/Lif1 binding. *DNA repair*. 2014;24:37-45.
- Chovanec M, Wilson TE. Restricting the ligation step of non-homologous end-joining. *DNA repair*. 2007;6(12):1890-1893.
- Chuang H-Y, Hofree M, Ideker T. A Decade of Systems Biology. *Annual Review of Cell and Developmental Biology*. 2010;26:721-744.
- Ciccia A, Elledge SJ. The DNA Damage Response: Making it safe to play with knives. *Molecular Cell*. 2010;40(2):179-204.
- Clemenson C, Marsolier-Kergoat MC. DNA damage checkpoint inactivation: Adaptation and recovery. *DNA Repair*. 2009;8:1101-1109.
- Close DM, Nelson WH, Bernhard WA. DNA damage by the direct effect of ionizing radiation: products produced by two sequential one-electron oxidations. *Journal of Physical Chemistry A*. 2013;117(47):12608-15.
- Conesa A, Mortazavi A. The common ground of genomics and systems biology. *BMC Systems Biology*. 2014;8(Suppl 2):S1.
- Costanzo M, Baryshnikova A, Myers CL, Andrews B, Boone C. Charting the genetic interaction map of a cell. *Current Opinion in Biotechnology*. 2011;22(1):66-74.
- Dai Y, Kysela B, Hanakahi LA, et al. Nonhomologous end joining and V(D)J recombination require an additional factor. *PNAS*. 2003;100(5):2462-2467.
- Daley JM, Kwon Y, Niu H, Sung P. Investigations of homologous recombination pathways and their regulation. *The Yale Journal of Biology and Medicine*. 2013;86(4):453-461.
- Daley JM, Niu H, Miller AS, Sung P. Biochemical mechanism of DSB end resection and its regulation. *DNA repair*. 2015;32: 66-74.

- Daley JM, Palmboos PL, Wu D, Wilson TE. Nonhomologous end joining in yeast. *Annu Rev Genet.* 2005;39:431-51.
- Davidson MB, Brown GW. Disecting the DNA damage response using functional genomics approaches in *S.cerevisiae*. *DNA Repair.* 2008;8:1110-1117.
- Davis DJ and Chen AJ. DNA double strand break repair via non-homologous end joining. *Translational Cancer Research.* 2013;2(3):130-143.
- De Bont R, van Larebeke N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis.* 2004;19(3):169-85.
- De Las Rivas J, Fontanillo C. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. Lewitter F, ed. *PLoS Computational Biology.* 2010;6(6):e1000807.
- Della Latta V, Cecchetti A, Del Ry S, Morales MA. Bleomycin in the setting of lung fibrosis induction: From biological mechanisms to counteractions. *Pharmacology Research.* 2015;97:122-30
- DeMille D, Grose JH. PAS Kinase: a nutrient sensing regulator of glucose homeostasis. *IUBMB life.* 2013;65(11):921-929.
- Deminoff SJ, Howard SC, Hester A, Warner S, Herman PK. Using Substrate-Binding Variants of the cAMP-Dependent Protein Kinase to Identify Novel Targets and a Kinase Domain Important for Substrate Interactions in *Saccharomyces cerevisiae*. *Genetics.* 2006;173(4):1909-1917.
- Deriano L and Roth DB. Modernizing the Nonhomologous End-Joining Repertoire: Alternative and Classical NHEJ Share the Stage. *Annual Review of Genetics.* 2013;47:433-455
- Deshpande RA, Wilson TE. Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. *DNA repair.* 2007;6(10):1507-1516.
- Dietlein F, Thelen L, Reinhardt CH. Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. *Trends in Genetics.* 2014;30(8):326–339.
- Dittrich MT, Klau GW, Rosenwald A, Dandekar T, Müller T. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. *Bioinformatics.* 2008;24(13):i223-i231.
- Do K, Chen AP. Molecular Pathways: Targeting PARP in Cancer Treatment. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2013;19(5):977-984.
- Downs JA, Allard S, Jobin-Robitaille O, Javaheri A, Auger A, Bouchard N, Kron SJ, Jackson SP, Côté J. Binding of Chromatin-Modifying Activities to Phosphorylated Histone H2A at DNA Damage Sites. *Molecular Cell.* 2004;16(6):979 – 990

Dungl DA, Maginn EN, Stronach EA. Preventing Damage Limitation: Targeting DNA-PKcs and DNA Double-Strand Break Repair Pathways for Ovarian Cancer Therapy. *Frontiers in Oncology*. 2015;5:240.

Elsztein C, de Lucena RM, de Morais MA. The resistance of the yeast *Saccharomyces cerevisiae* to the biocide polyhexamethylene biguanide: involvement of cell wall integrity pathway and emerging role for YAP1. *BMC Molecular Biology*. 2011;12:38.

Enns LC, Pettan-Brewer C, Ladiges W. Protein kinase A is a target for aging and the aging heart. *Aging (Albany NY)*. 2010;2(4):238-243.

Faucher D, Wellinger RJ. Methylated H3K4, a Transcription-Associated Histone Modification, Is Involved in the DNA Damage Response Pathway. Copenhagen GP, ed. *PLoS Genetics*. 2010;6(8):e1001082.

Ferretti LP, Lafranchi L, Sartori AA. Controlling DNA-end resection: a new task for CDKs. *Frontiers in Genetics*. 2013;4:99.

Fiedler D, Braberg H, Mehta M, et al. Functional Organization of the *S. cerevisiae* Phosphorylation Network. *Cell*. 2009;136(5):952-963.

Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature*. 1989; 340:245-6.

Finn K, Lowndes NF, Grenon M. Eukaryotic DNA damage checkpoint activation in response to double-strand breaks. *Cellular and Molecular Life Sciences*. 2012;69(9):1447-73.

Finn K, Lowndes NF, Grenon M. Eukaryotic DNA damage checkpoint activation in response to double-strand breaks. *Cellular and Molecular Life Sciences*. 2012;69(9):1447-73.

Forler D, Kocher T, Rode M, et al. An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nature Biotechnology*. 2003;21:89-92.

Fortin GS, Symington LS. Mutations in yeast Rad51 that partially bypass the requirement for Rad55 and Rad57 in DNA repair by increasing the stability of Rad51–DNA complexes. *The EMBO Journal*. 2002;21(12):3160-3170.

Friedberg EC, McDaniel LD, Schultz RA. The role of endogenous and exogenous DNA damage and mutagenesis. *Current Opinion in Genetics and Development*. 2004;14(1):5-10.

Frit P, Barboule N, Yuan Y, Gomez D, Calsou P. Alternative end-joining pathway(s): bricolage at DNA breaks. *DNA Repair Amst*. 2014;17:81-97.

Galello F, Portela P, Moreno S, Rossi S. Characterization of Substrates That Have a Differential Effect on *Saccharomyces cerevisiae* Protein Kinase A Holoenzyme Activation. *The Journal of Biological Chemistry*. 2010;285(39):29770-29779.

Galván IJ, Mir-Rashed N, Jessulat M, et al. Antifungal and antioxidant activities of the phytomedicine pipsissewa, *Chimaphila umbellata*. *Phytochemistry*. 2008;69(3):738-46.

- 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):108-12.
- Garcia V, Phelps SEL, Gray S, Neale MJ. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature*. 2011;479(7372):241-244.
- Ghodke I, Muniyappa K. Processing of DNA Double-stranded Breaks and Intermediates of Recombination and Repair by *Saccharomyces cerevisiae* Mre11 and Its Stimulation by Rad50, Xrs2, and Sae2 Proteins . *The Journal of Biological Chemistry*. 2013;288(16):11273-11286.
- Gomez SM, Noble WS, Rzhetsky A. Learning to predict protein-protein interactions from rotein sequences. *Bioinformatics*. 2003;19(15):1875-81.
- Grose J.H, Smith T.L, Sabic H, Rutter J. Yeast PAS kinase coordinates glucose partitioning in response to metabolic and cell integrity signaling. *EMBO*. 2007;28:4824-30.
- Grose JH, Smith TL, Sabic H, Rutter J. Yeast PAS kinase coordinates glucose partitioning in response to metabolic and cell integrity signaling. *The EMBO journal*. 2007 Nov 28;26(23):4824-30.
- Guénolé A, Srivas R, Vreeken K, et al. Dissection of DNA Damage Responses Using Multiconditional Genetic Interaction Maps. *Molecular cell*. 2013;49(2):346-358.
- 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 Research*. 2008;36(9):3025-3030.
- Hamilton G, Yee KS, Scrace S, O’Neill E. ATM Regulates a RASSF1A-Dependent DNA Damage Response. *Current Biology*. 2009;19:2020–2025.
- Hefferin ML, Tomkinson AE. Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair*. 2005;4:639–648.
- Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nature Review Genetics*. 2014;15(9):585-98.
- Hendry JA, Tan G, Ou J, Boone C, Brown GW. Leveraging DNA Damage Response Signaling to Identify Yeast Genes Controlling Genome Stability. *G3: Genes|Genomes|Genetics*. 2015;5(5):997-1006.
- Hentges P, Waller H, Reis CC, Ferreira MG, Doherty AJ. Cdk1 Restrains NHEJ through Phosphorylation of XRCC4-like Factor Xlf1. *Cell Reports*. 2014;9(6):2011-2017.
- Heo J, Li J, Summerlin M, Hays A, et al. TDP1 promotes assembly of non-homologous end joining protein complexes on DNA. *DNA Repair (Amst)*. 2015;30:28-37.

- Hopfner KP, Karchner A, Shin DS et al. Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell*. 2000;101:798–800
- Horigome C, Oma Y, Konishi T, Schmid R, Marcomini I, Hauer MH, Dion V, Harata M, Gasser SM. SWR1 and INO80 chromatin remodelers contribute to DNA double-strand break perinuclear anchorage site choice. *Molecular Cell*. 2014;55(4):626-39.
- Hsu F-M, Zhang S, Chen BPC. Role of DNA-dependent protein kinase catalytic subunit in cancer development and treatment. *Translational cancer research*. 2012;1(1):22-34.
- Hu P, Janga SC, Babu M, et al. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biology*. 2009;7(4):e1000096.
- Huang M, Xu Q, Mitsui K, Xu Z. PSK1 regulates expression of SOD1 involved in oxidative stress tolerance in yeast. *FEMS microbiology letters*. 2014;350(2):154-60.
- Huertas P, Cortés-Ledesma F, Sartori AA, Aguilera A, Jackson SP. CDK targets Sae2 to control DNA-end resection and homologous recombination. *Nature*. 2008;455(7213):689-692.
- Ikehata H, Ono T. The mechanisms of UV mutagenesis. *Journal of Radiation Research*. 2011;52(2):115-25.
- Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature*. 2009;461(7267):1071-1078.
- Jackson SP. Sensing and repairing DNA double-strand breaks. *Carcinogenesis*. 2002;23:687-696.
- Jäger S, Cimermancic P, Gulbahce N, et al. Global landscape of HIV–human protein complexes. *Nature*. 2011;481(7381):365-370.
- Jain E, Bairoch A, Duvaud S, et al. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*. 2009;10:136.
- Jasin M, de Villiers J, Weber F, Schaffner W. High frequency of homologous recombination in mammalian cells between endogenous and introduced SV40 genomes. *Cell*. 1985;43(3 Pt 2):695-703.
- Jasin M, Rothstein R. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harbor Perspectives in Biology*. 2013;5(11):a012740.
- Jekimovs C, Bolderson E, Suraweera A, Adams M, O’Byrne KJ, Richard DJ. Chemotherapeutic Compounds Targeting the DNA Double-Strand Break Repair Pathways: The Good, the Bad, and the Promising. *Frontiers in Oncology*. 2014;4:86.

- Jelinsky SA, Estep P, Church GM, Samson LD. Regulatory Networks Revealed by Transcriptional Profiling of Damaged *Saccharomyces cerevisiae* Cells: Rpn4 Links Base Excision Repair with Proteasomes. *Molecular and Cellular Biology*. 2000;20(21):8157-8167.
- Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001 ;411(6833):41-2.
- 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*. *Archive Biochem Biophys*. 2008;469(2):157-64.
- Jessulat M, Buist T, Alamgir M, et al. In vivo investigation of protein-protein interactions for helicases using tandem affinity purification. *Methods Molecular Biology*. 2010;587:99-111.
- Jessulat M, Maltby RH, Nguyen-Tran DH, et al. Spindle checkpoint factors Bub1 and Bub2 promote DNA double-strand break repair by nonhomologous end joining. *Molecular Cell Biology*. 2015;35(14):2448-63.
- Jessulat M, Pitre S, Gui Y, et al. Recent advances in protein-protein interaction prediction: experimental and computational methods. *Expert Opinion Drug Discovery*. 2011;6(9):921-35.
- Kapoor P, Shen X. Mechanisms of Nuclear Actin in Chromatin Remodeling Complexes. *Trends in cell biology*. 2014;24(4):238-246.
- Kelley MR, Logsdon D, Fishel ML. Targeting DNA repair pathways for cancer treatment: what's new? *Future oncology (London, England)*. 2014;10(7):1215-1237.
- Kerppola TK. Complementary Methods for Studies of Protein Interactions in Living Cells. *Nature methods*. 2006;3(12):969-971.
- Keskin O, Nussinov R, Gursoy A. PRISM: Protein-protein Interaction Prediction by Structural Matching. *Methods in molecular biology (Clifton, NJ)*. 2008;484:505-521.
- Klutstein M, Siegfried Z, Gispan A, et al. Combination of genomic approaches with functional genetic experiments reveals two modes of repression of yeast middle-phase meiosis genes. *BMC Genomics*. 2010;11:478.
- Knuesel M, Wan Y, Xiao Z, et al. Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. *Molecular Cell Proteomics*. 2003;2:1225-33.
- Koç A, Wheeler LJ, Mathews CK, Merrill GF. Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. *Journal Biological Chemistry*. 2004;279(1):223-30.
- Koh JLY, Ding H, Costanzo M, et al. DRYGIN: a database of quantitative genetic interaction networks in yeast. *Nucleic Acids Research*. 2010;38(Database issue):D502-D507.

- Krejci L, Altmannova V, Spirek M, Zhao X. Homologous recombination and its regulation. *Nucleic Acids Research*. 2012;40(13):5795-5818.
- Krogan NJ, Cagney G, Yu H, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*. 2006;440(7084):637-43.
- Kuzmin E, Sharifpoor S, Baryshnikova A, et al. Synthetic genetic array analysis for global mapping of genetic networks in yeast. *Methods in Molecular Biology*. 2014;1205:143-68.
- Larsen DH, Poinsignon C, Gudjonsson T, et al. The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage. *The Journal of Cell Biology*. 2010;190(5):731-740.
- Lee K, Lee SE. *Saccharomyces cerevisiae* Sae2- and Tel1-Dependent Single-Strand DNA Formation at DNA Break Promotes Microhomology-Mediated End Joining. *Genetics*. 2007;176(4):2003-2014.
- Lee SE, Moore JK, Holmes A, Umezumi K, Kolodner RD, Haber JE. *Saccharomyces* Ku70, mre11/rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell*. 1998;94(3):399-409.
- Li S, Wang L, Fu B, Berman MA, Diallo A, Dorf ME. TRIM65 regulates microRNA activity by ubiquitination of TNRC6. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(19):6970-6975.
- Liang B, Qiu J, Ratnakumar K, Laurent BC. RSC functions as an early double-strand break sensor in the cell's response to DNA damage. *Current biology : CB*. 2007;17(16):1432-1437.
- Lieber MR. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annual review of biochemistry*. 2010;79:181-211.
- Liu J, Jia Y, Li J, Chu Z. Transcriptional profiling analysis of individual kinase-deletion strains of fission yeast in response to nitrogen starvation. *Molecular Genetics Genomics*. 2015;290(3):1067-83.
- Liu P, Gan W, Guo C, et al. Akt-mediated phosphorylation of XLF impairs non-homologous end-joining DNA repair. *Molecular Cell*. 2015 Feb 19;57(4):648-61.
- Liu X, Jiang W, Dubois RL, Yamamoto K, Wolner Z, Zha S. Overlapping functions between XLF repair protein and 53BP1 DNA damage response factor in end joining and lymphocyte development. *PNAS*. 2012;109(10):3903-3908.
- Liu ZP, Chen L. Proteome-wide prediction of protein-protein interactions from high-throughput data. *Protein Cell*. 2012;3(7):508-20.
- Magtanong L, Ho CH, Barker SL, et al. Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nature Biotechnology*. 2011;29(6):505-11.

- Mahaney BL, Lees-Miller SP, Cobb JA. The C-terminus of Nej1 is critical for nuclear localization and non-homologous end-joining. *DNA repair*. 2014;14:9-16.
- Mahaney BL, Meek K, Lees-Miller SP. Repair of ionizing radiation-induced DNA double strand breaks by non-homologous end-joining. *The Biochemical journal*. 2009;417(3):639-650.
- Makhnevych T, Sydorskyy Y, Xin X, et al. Raught B. Global map of SUMO function revealed by protein-protein interaction and genetic networks. *Molecular Cell*. 2009;33(1):124-35.
- Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. *Science*. 1999;285(5428):751-3.
- Maréchal A, Zou L. DNA Damage Sensing by the ATM and ATR Kinases. *Cold Spring Harbor Perspectives in Biology*. 2013;5(9):a012716.
- Martin S, Roe D, Faulon JL. Predicting protein-protein interactions using signature products. *Bioinformatics*. 2005;21(2):218-26
- Matsuzaki K, Shinohara A, Shinohara M. Forkhead-Associated Domain of Yeast Xrs2, a Homolog of Human Nbs1, Promotes Nonhomologous End Joining Through Interaction With a Ligase IV Partner Protein, Lif1. *Genetics*. 2008;179(1):213-225.
- McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genetics*. 2008;24(11):529-38.
- Mehta A, Haber JE. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harbor Perspective in Biology*. 2014;6(9):a016428.
- Memarian N, Jessulat M, Alirezaie J, et al. Colony size measurement of the yeast gene deletion strains for functional genomics. *BMC Bioinformatics*. 2007;8:117.
- Mimitou EP, Symington LS. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*. 2008;455(7214):10.1038/nature07312.
- Miura T, Yamana Y, Usui T, Ogawa HI, Yamamoto M-T, Kusano K. Homologous Recombination via Synthesis-Dependent Strand Annealing in Yeast Requires the Irc20 and Srs2 DNA Helicases. *Genetics*. 2012;191(1):65-78.
- Modesti M, Kanaar R. Homologous recombination: from model organisms to human disease. *Genome Biology*. 2001;2(5):1014.
- Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*. 1996;16(5):2164-2173.
- Morrison AJ, Shen X. Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nature Review Molecular Cell Biology*. 2009;10(6):373-84.

- Muley VY, Ranjan A. Evaluation of Physical and Functional Protein-Protein Interaction Prediction Methods for Detecting Biological Pathways. Ouzounis CA, ed. *PLoS ONE*. 2013;8(1):e54325.
- Nakada D, Matsumoto K, and Sugimoto K. ATM-related Tel1 associates with double-strand breaks through Xrs2-dependent mechanism. *Genes and Development*. 2003;17:1957-1962
- Nassa G, Tarallo R, Ambrosino C, et al. A large set of estrogen receptor β -interacting proteins identified by tandem affinity purification in hormone-responsive human breast cancer cell nuclei. *Proteomics*. 2011;11:159-65.
- Neduva V, Linding R, Su-Angrand I, et al. Systematic Discovery of New Recognition Peptides Mediating Protein Interaction Networks. Matthews R, ed. *PLoS Biology*. 2005;3(12):e405.
- Novarina D, Amara F, Lazzaro F, Plevani P, Muzi-Falconi M. Mind the gap: Keeping UV lesions in check. *DNA Repair*. 2011;10(7):751-759.
- O'Connor MJ. Targeting the DNA Damage Response in Cancer. *Molecular Cell*. 2015;60:547-560
- Ohue M, Matsuzaki Y, Shimoda T, Ishida T, Akiyama Y. Highly precise protein-protein interaction prediction based on consensus between template-based and *de novo* docking methods. *BMC Proceedings*. 2013;7(Suppl 7):S6.
- Omidi K, Hooshyar M, Jessulat M, et al. Phosphatase Complex Pph3/Psy2 Is Involved in Regulation of Efficient Non-Homologous End-Joining Pathway in the Yeast *Saccharomyces cerevisiae*. Cotterill S, ed. *PLoS ONE*. 2014;9(1):e87248.
- Ooi SL, Shoemaker DD, Boeke JD. A DNA microarray-based genetic screen for nonhomologous end-joining mutants in *Saccharomyces cerevisiae*. *Science*. 2001;294(5551):2552
- Palmbos PL, Wu D, Daley JM, Wilson TE. Recruitment of *Saccharomyces cerevisiae* Dnl4–Lif1 Complex to a Double-Strand Break Requires Interactions With Yku80 and the Xrs2 FHA Domain. *Genetics*. 2008;180(4):1809-1819.
- Pan MR, Hsieh HJ, Dai H, et al. Chromodomain Helicase DNA-binding Protein 4 (CHD4) Regulates Homologous Recombination DNA Repair, and Its Deficiency Sensitizes Cells to Poly(ADP-ribose) Polymerase (PARP) Inhibitor Treatment. *The Journal of Biological Chemistry*. 2012;287(9):6764-6772.
- Pan X, Lei B, Zhou N, et al. Identification of novel genes involved in DNA damage response by screening a genome-wide *Schizosaccharomyces pombe* deletion library. *BMC Genomics*. 2012;13:662.
- Pannunzio NR, Li S, Watanabe G, Lieber MR. NHEJ Often Uses Microhomology: Implications for Alternative End Joining. *DNA repair*. 2014;17:74-80.

- Park Y. Critical assessment of sequence-based protein-protein interaction prediction methods that do not require homologous protein sequences. *BMC Bioinformatics*. 2009;10:419.
- Parsons AB, Lopez A, Givoni IE, et al. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell*. 2006;126(3):611-625.
- Pearl LH, Schierz AC, Ward SE, Al-Lazikani B, Pearl FM. Therapeutic opportunities within the DNA damage response. *Nature Review Cancer*. 2015;15(3):166-80.
- Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein-protein interactions: structural, functional, and network properties. *Structure*. 2010;18(10):1233-43.
- Petukhova G, Stratton S, Sung P. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature*. 1998;393(6680):91-4.
- Pitre S, Alamgir M, Green JR, Dumontier M, Dehne F, Golshani A. Computational methods for predicting protein-protein interactions. *Adv Biochem Eng Biotechnol*. 2008a;110:247-67.
- Pitre S, Dehne F, Chan A, et al. 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.
- Pitre S, Hooshyar M, Schoenrock A, et al. Short Co-occurring Polypeptide Regions Can Predict Global Protein Interaction Maps. *Scientific Reports*. 2012;2:239.
- Pitre S, North C, Alamgir M, et al. Global investigation of protein-protein interactions in yeast *Saccharomyces cerevisiae* using re-occurring short polypeptide sequences. *Nucleic Acids Research*. 2008b;36(13):4286-4294.
- Povirk LF, Hogan M, Dattagupta N, Buechner M. Copper(II).bleomycin, iron(III).bleomycin, and copper(II).phleomycin: comparative study of deoxyribonucleic acid binding. *Biochemistry*. 1981;20(3):665-71.
- Powell SN, Bindra RS. Targeting the DNA damage response for cancer therapy. *DNA Repair (Amst)*. 2009;2;8(9):1153-65.
- Price BD, D'Andrea AD. Chromatin Remodeling at DNA Double Strand Breaks. *Cell*. 2013;152(6):1344-1354.
- Ptacek J, Devgan G, Michaud G, et al. Global analysis of protein phosphorylation in yeast. *Nature*. 2005;438(7068):679-84.
- Puig O, Caspary F, Rigaut G, et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*. 2001;24:218-29.
- Putman CD, Jaechnig EC, Kolodner RD. Prespectives on the DNA damage and replication responses in *Saccharomyces cerevisiae*. *DNA Repair*. 2009;8:974-982.
- Rass U, Ahel I, West SC. Defective DNA Repair and Neurodegenerative Disease. *Cell*. 2007;130(6):991-1004.

Richardson E, Stockwell SR, Li H, Aherne W, Cuomo ME, Mittnacht S. Mechanism-based screen establishes signalling framework for DNA damage-associated G1 checkpoint response. Cotterill S, ed. *PLoS ONE*. 2012;7(2):e31627.

Roy S, Andres SN, Vergnes A, et al. XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Research*. 2012;40(4):1684-1694.

Ruiz JF, Pardo B, Sastre-Moreno G, Aguilera A, Blanco L. Yeast Pol4 Promotes Tel1-Regulated Chromosomal Translocations. Jinks-Robertson S, ed. *PLoS Genetics*. 2013;9(7):e1003656.

Saleh-Gohari N, Bryant HE, Schultz N, Parker KM, Cassel TN, Helleday T. Spontaneous Homologous Recombination Is Induced by Collapsed Replication Forks That Are Caused by Endogenous DNA Single-Strand Breaks. *Molecular and Cellular Biology*. 2005;25(16):7158-7169.

Salton M, Lerenthal Y, Wang SY, Chen DJ, Shiloh Y. Involvement of Matrin 3 and FPQ/NONO in the DNA damage response. *Cell Cycle*. 2010;9(8):1568-76.

Samanfar B, Omid K, Hooshyar M, et al. Large-scale investigation of oxygen response mutants in *Saccharomyces cerevisiae*. *Molecular Biosystem*. 2013;9(6):1351-9.

San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem*. 2008;77:229-57.

Schipler A, Iliakis G. DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. *Nucleic Acids Research*. 2013;41(16):7589-7605.

Schoenrock A, Samanfar B, Pitre S, et al. Efficient prediction of human protein-protein interactions at a global scale. *BMC Bioinformatics*. 2014;15(1):383.

Searle JS, Schollaert KL, Wilkins BJ, Sanchez Y. The DNA damage checkpoint and PKA pathways converge on APC substrates and Cdc20 to regulate mitotic progression. *Nature Cell Biology*. 2004;6(2):138-45.

Searle JS, Wood MD, Kaur M, Tobin DV, Sanchez Y. Proteins in the Nutrient-Sensing and DNA Damage Checkpoint Pathways Cooperate to Restrain Mitotic Progression following DNA Damage. Cohen-Fix O, ed. *PLoS Genetics*. 2011;7(7):e1002176.

Seluanov A, Mittelman D, Pereira-Smith OM, Wilson JH, Gorbunova V. DNA end joining becomes less efficient and more error-prone during cellular senescence. *PNAS*. 2004;101(20):7624-7629.

Shalem O, Sanjana NE, Hartenian E, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science (New York, NY)*. 2014;343(6166):84-87.

Sharifpoor S, van Dyk D, Costanzo M, et al. Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs. *Genome Research*. 2012;22(4):791-801.

- Shen J, Zhang J, Luo X, et al. Predicting protein–protein interactions based only on sequences information. *PNAS*. 2007;104(11):4337-4341.
- Shibata A, Jeggo PA. DNA double-strand break repair in a cellular context. *Clinical Oncology*. 2014;26:243e249.
- Shim EY, Hong SJ, Oum J-H, Yanez Y, Zhang Y, Lee SE. RSC Mobilizes Nucleosomes To Improve Accessibility of Repair Machinery to the Damaged Chromatin . *Molecular and Cellular Biology*. 2007;27(5):1602-1613.
- Shim EY, Ma J-L, Oum J-H, Yanez Y, Lee SE. The yeast chromatin remodeler RSC complex facilitates end joining repair of DNA double-strand breaks. *Molecular and Cellular Biology*. 2005;25(10):3934-3944.
- Shrivastav M, P De Haro LP, and Nickoloff J.A. Regulation of DNA double-strand break repair pathway choice. *Cell Research*. 2008;18:134-147.
- Shrivastav N, Li D, Essigmann JM. Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis*. 2010;31(1):59-70.
- Sinha M, Peterson CL. Chromatin dynamics during repair of chromosomal DNA double strand breaks. *Epigenomics*. 2009;1(2): 371–385.
- Skoneczna A, Kaniak A, Skoneczny M. Genetic instability in budding and fission yeast sources and mechanisms. Danchin A, ed. *FEMS Microbiology Reviews*. 2015;39(6):917-967.
- Skoneczna A, Kaniak A, Skoneczny M. Genetic instability in budding and fission yeast—sources and mechanisms. *FEMS Microbiology Review*. 2015; 39(6): 917–967.
- Skrabanek L, Saini HK, Bader GD *et al.* Computational prediction of protein-protein interactions. *Molecular Biotechnology*. 2008;38:1-17.
- Sobanski T, Fraporti L, Almeida APD. Proteins Involved in DNA Double-Strand Breaks Repair Pathways Are Essential to Prevent the Development of Cancer. *Journal of Cancer Prevention Current Research*. 2015;3(3): 00081.
- Sopko R, Andrews BJ. Linking the kinome and phosphorylome--a comprehensive review of approaches to find kinase targets. *Molecular Biosystem*. 2008;4(9):920-33.
- Sopko R, Huang D, Preston N, et al. Mapping pathways and phenotypes by systematic gene overexpression. *Molecular Cell*. 2006;21(3):319-30.
- Srivastava M, Nambiar M, Sharma S, et al. An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell*. 2012;151:1474-1487.
- Stotz HU, Findling S, Nukarinen E, Weckwerth W, Mueller MJ, Berger S. A tandem affinity purification tag of TGA2 for isolation of interacting proteins in *Arabidopsis thaliana*. *Plant Signaling & Behavior*. 2014;9(10):e972794.

- Stracker TH, Usui T, Petrini JHJ. Taking the time to make important decisions: the checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. *DNA Repair*. 2009;8(9):1047-1054.
- Sugiyama T, Kantake N, Wu Y, Kowalczykowski SC. Rad52-mediated DNA annealing after Rad51-mediated DNA strand exchange promotes second ssDNA capture. *EMBO*. 2006;25(23):5539-5548.
- Sugiyama T, Kantake N. Dynamic Regulatory Interactions of Rad51, Rad52, and Replication Protein-A in Recombination Intermediates, *Journal Molecular Biology*. 2009;390(1):45-55.
- Sugiyama T, Zaitseva EM, Kowalczykowski SC. A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *Journal Biological Chemistry*. 1997;272(12):7940-5.
- Suter B, Kittanakom S, Stagljar I. Two-hybrid technologies in proteomics research. *Current Opinion Biotechnology*. 2008;19:316-23.
- Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. *Annual Review Genetics*. 2011;45:247-71.
- Symington LS. Role of *RAD52* Epistasis Group Genes in Homologous Recombination and Double-Strand Break Repair. *Microbiology and Molecular Biology Reviews*. 2002;66(4):630-670.
- Szerlong H, Saha A, Cairns BR. The nuclear actin-related proteins Arp7 and Arp9: a dimeric module that cooperates with architectural proteins for chromatin remodeling. *EMBO*. 2003;22(12):3175-3187.
- Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. *Cell*. 1983;33(1):25-35.
- Teo SH, Jackson SP. Lif1p targets the DNA ligase Lig4p to sites of DNA double-strand breaks. *Curr Biol*. 2000;10(3):165-168.
- Thompson BJ and Sahai E. MST kinases in development and disease. *Journal of Cell Biology*. 2015;210(6): 871-882.
- Tong AH, Boone C. Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Molecular Biology*. 2006;313:171-92.
- Tosi A, Haas C, Herzog F, et al. Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell*. 2013;154(6):1207-19.
- Tosti E, Katakowski JA, Schaezlein S, et al. Evolutionarily conserved genetic interactions with budding and fission yeast *MutS* identify orthologous relationships in mismatch repair-deficient cancer cells. *Genome Medicine*. 2014;6(9):68.

- Trujillo KM, Roh DH, Chen L, Van Komen S, Tomkinson A, Sung P. Yeast xrs2 binds DNA and helps target rad50 and mre11 to DNA ends. *Journal of Biological Chemistry*. 2003;278(49):48957-64.
- Tsukuda T, Fleming AB, Nickoloff JA, Ann Osley M. Chromatin Remodeling at a DNA Double-Strand Break Site in *Saccharomyces cerevisiae*. *Nature*. 2005;438(7066):379-383.
- Van Komen S, Petukhova G, Sigurdsson S, Stratton S, Sung P. Superhelicity-driven homologous DNA pairing by yeast recombination factors Rad51 and Rad54. *Molecular Cell*. 2000;6(3):563-72.
- Van Zuylen WJ, Doyon P, Clément J-F, et al. Proteomic Profiling of the TRAF3 Interactome Network Reveals a New Role for the ER-to-Golgi Transport Compartments in Innate Immunity. Thiel V, ed. *PLoS Pathogens*. 2012;8(7):e1002747.
- Villarreal DD, Lee K, Deem A, Shim EY, Malkova A, Lee SE. Microhomology Directs Diverse DNA Break Repair Pathways and Chromosomal Translocations. Lichten M, ed. *PLoS Genetics*. 2012;8(11):e1003026.
- Weterings E1, Chen DJ. The endless tale of non-homologous end-joining. *Cell Research*. 2008;18(1):114-24.
- Wiegman AP, Al-Ejeh F, Chee N, et al. Rad51 supports triple negative breast cancer metastasis. *Oncotarget*. 2014;5(10):3261-3272.
- Williams RS, Moncalian G, Williams JS, et al. Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair. *Cell*. 2008;135(1):97-109.
- Winzler EA1, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*. 1999;285(5429):901-6.
- Woodbine L, Gennery AR, Jeggo PA. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair*. 2014;16:84–96.
- Wu D, Topper LM, Wilson TE. Recruitment and Dissociation of Nonhomologous End Joining Proteins at a DNA Double-Strand Break in *Saccharomyces cerevisiae*. *Genetics*. 2008;178(3):1237-1249.
- Wu WH, Alami S, Luk E, Wu CH, Sen S, Mizuguchi G, Wei D, Wu C. Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nature Structural Molecular Biology*. 2005; 12(12):1064-71.
- Wu ZL. Phosphatase-Coupled Universal Kinase Assay and Kinetics for First-Order-Rate Coupling Reaction. Uversky VN, ed. *PLoS ONE*. 2011;6(8):e23172.
- Xu X, Song Y, Li Y, Chang J, Zhang H, An L. The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. *Protein Expression and Purification*. 2010;72(2):149-56.

- Yang H, Matsumoto Y, Trujillo KM, Lees-Miller SP, Osley MA, Tomkinson AE. Role of the yeast DNA repair protein Nej1 in end processing during the repair of DNA double strand breaks by non-homologous end joining. *DNA Repair*. 2015;31:1-10.
- Yang W. Surviving the sun: Repair and bypass of DNA UV lesions. *Protein Science : A Publication of the Protein Society*. 2011;20(11):1781-1789.
- Yano K, Morotomi-Yano K, Wang S-Y, et al. Ku recruits XLF to DNA double-strand breaks. *EMBO Reports*. 2008;9(1):91-96.
- Yoko-o T, Roy SK, Jigami Y. Differences in in vivo acceptor specificity of two galactosyltransferases, the gmh3+ and gma12+ gene products from *Schizosaccharomyces pombe*. *Eur J Biochem*. 1998;257(3):630-7.
- Yoshida T, Shimada K, Oma Y, et al. Actin-Related Protein Arp6 Influences H2A.Z-Dependent and -Independent Gene Expression and Links Ribosomal Protein Genes to Nuclear Pores. Bickmore WA, ed. *PLoS Genetics*. 2010;6(4):e1000910.
- You Z-H, Zhu L, Zheng C-H, Yu H-J, Deng S-P, Ji Z. Prediction of protein-protein interactions from amino acid sequences using a novel multi-scale continuous and discontinuous feature set. *BMC Bioinformatics*. 2014;15(Suppl 15):S9.
- Yu CY, Chou LC, Chang DTH. Predicting protein-protein interactions in unbalanced data using the primary structure of proteins. *BMC Bioinformatics*. 2010;11:167.
- Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. Murray D, ed. *PLoS Computational Biology*. 2007;3(4):e59.
- Yu L, Volkert MR. Differential requirement for *SUB1* in chromosomal and plasmid double-strand DNA break repair. Lichten M, ed. *PLoS ONE*. 2013;8(3):e58015.
- Yuan F, Xie Q, Wu J, et al. MST1 Promotes apoptosis through regulating Sirt1-dependent p53 deacetylation. *The Journal of Biological Chemistry*. 2011;286(9):6940-6945.
- Zahiri J, Bozorgmehr JH, Masoudi-Nejad A. Computational prediction of protein-protein interaction networks: Algo-rithms and Resources. *Current Genomics*. 2013;14(6):397-414.
- Zaki N, Lazarova-Molnar S, El-Hajj W, Campbell P. Protein-protein interaction based on pairwise similarity. *BMC Bioinformatics*. 2009;10:150.
- Zeman MK, Cimprich KA. Causes and consequences of replication stress. *Nature cell biology*. 2014;16(1):2-9.
- Zha S, Jiang W, Fujiwara Y, et al. Ataxia telangiectasia-mutated protein and DNA-dependent protein kinase have complementary V(D)J recombination functions. *PNAS*. 2011;108(5):2028-2033.

Zhang Y, Hefferin ML, Chen L, et al. Role of Dnl4-Lif1 in nonhomologous end-joining repair complex assembly and suppression of homologous recombination. *Nature Structural Molecular Biology*. 2007;14(7):639- 646.

Zhu Z, Chung W-H, Shim EY, Lee SE, Ira G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double strand break ends. *Cell*. 2008;134(6):981-994.

Zou L and Elledge SJ. Sensing the damage through ATRIP recognition of RPA-ssDNA complexes. *Science*. 2003;300:1542-1548.

Appendix 1

List of publications

- Jessulat M, Malty RH, Nguyen-Tran DH, Deineko V, Aoki H, Vlasblom J, Omid K, Jin K, Minic Z, **Hooshyar M**, Burnside D, Samanfar B, Phanse S, Freywald T, Prasad B, Zhang Z, Vizeacoumar F, Krogan NJ, Freywald A, Golshani A, Babu M. Spindle Checkpoint Factors Bub1 and Bub2 Promote DNA Double-Strand Break Repair by Nonhomologous End Joining. *Mol Cell Biol*. 2015. 35(14):2448-63
- Schoenrock A, Samanfar B, Pitre S, **Hooshyar M**, Jin K, Phillips CA, Wang H, Phanse S, Omid K, Gui Y, Alamgir M, Wong A, Barrenäs F, Babu M, Benson M, Langston MA, Green JR, Dehne F, Golshani A. Efficient prediction of human protein-protein interactions at a global scale. *BMC Bioinformatics*. 2014. 5(1):383.
- Samanfar B, Tan L.H, Shostak K, Chalabian F, Wu Z, Alamgir M.D, Sunba N, Burnside D, Omid K, **Hooshyar M**, Galván Márquez I, Babu M and Golshani A: A global investigation of gene deletion strains that affect premature stop codon bypass in yeast, *Saccharomyces cerevisiae*. *Molecular biosystems*. 2014. 10(4):916-24.
- 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.
- 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*. *Molecular Biosystem*. 2013. 9(6):1351-1359
- Pitre S, **Hooshyar M**, Schoenrock A, Samanfar B, Jessulat M, Green J.R, Dehne F, Golshani A. Short co-occurring polypeptide regions can predict global protein interaction maps. *Scientific Report*. 2012. 2: 239.
- Jessulat M, Pitre S, Gui Y, **Hooshyar M**, Omid K, Samanfar B, Tan L.H, Alamgir M, Green J, Dehne F, Golshani A. Recent advances in protein-protein interaction prediction: experimental and computational methods. *Expert Opinion in Drug Discovery*. 2011. 6(9): 921-35.

- Jessulat M, Buist T, Alamgir M, **Hooshyar M**, Xu J, Aoki H, Ganoza M.C, Butland G, Golshani A. In vivo investigation of protein-protein interactions for helicases using tandem affinity purification. *Methods in Molecular Biology*. 2010. **587**: 99-111.

Appendix 2

Appendix 2.1 ordered list of PPIs predictions

Protein A	Protein B	Score	Status	Same component	Same function	Same process	Third Party Interaction
SPCC23B6.03e	SPBC216.05	0.9999994	Known	NO	YES	NO	YES
SPAC3G9.12	SPAC1093.06e	0.9999993	Known	YES	NO	NO	YES
SPBC216.07e	SPBC30D10.10e	0.9999993	Known	YES	YES	YES	YES
SPBC216.05	SPBC25H2.13e	0.9999992	Known	NO	NO	NO	YES
SPBC216.02	SPAC1093.06e	0.9999999	Known	YES	NO	YES	NO
SPAC4A8.05e	SPCC645.05e	0.9999987	Known	YES	YES	YES	YES
SPAC4F10.13e	SPAC19D5.04	0.9999987	Known	NO	NO	YES	NO
SPBP16F5.03e	SPAC458.03	0.9999987	Known	YES	YES	YES	NO
SPAC2G11.12	SPBC216.05	0.9999987	Known	NO	YES	YES	YES
SPBC19G7.05e	SPCC5E4.04	0.9999987	Known	NO	NO	NO	NO
SPAC458.03	SPAC1F5.11e	0.9999987	Known	NO	YES	NO	NO
SPAC630.13e	SPBC30D10.10e	0.9999987	Known	YES	NO	YES	YES
SPAC57A7.11	SPBC216.07e	0.9999987	Known	YES	YES	YES	YES
SPBC12C2.02e	SPBC30D10.10e	0.9999987	Known	YES	NO	YES	NO
SPCC5E4.04	SPAC31A2.05e	0.9999986	Known	NO	YES	NO	NO
SPCC23B6.03e	SPAC694.06e	0.9999986	Known	NO	YES	NO	YES
SPAC27D7.13e	SPAC1093.06e	0.9999985	Known	YES	NO	YES	YES
SPBC106.09	SPCC5E4.04	0.9999985	Known	NO	NO	NO	YES
SPBC3E7.08e	SPBC216.05	0.9999985	Known	NO	NO	NO	YES
SPBC336.04	SPBC216.05	0.9999984	Known	NO	YES	NO	YES
SPCC622.13e	SPBC30D10.10e	0.9999984	Known	NO	NO	NO	YES
SPBC216.07e	SPCC622.13e	0.9999984	Known	NO	YES	NO	YES
SPAC694.06e	SPBC216.05	0.9999983	Known	YES	YES	YES	YES
SPCC23B6.03e	SPAC458.03	0.9999983	Known	NO	YES	NO	YES
SPBC12C2.02e	SPCC5E4.04	0.9999983	Known	YES	NO	NO	NO
SPBC336.04	SPBC25H2.13e	0.9999983	Known	YES	YES	NO	YES
SPCC1620.14e	SPAC3G6.01	0.9999982	Known	YES	YES	NO	NO
SPBC29A10.04	SPCC5E4.04	0.9999982	Known	NO	YES	NO	YES
SPAC4A8.05e	SPCC1739.11e	0.9999981	Known	NO	YES	YES	YES
SPAC3F10.11e	SPAC30.04e	0.9999981	Known	YES	YES	YES	NO
SPBC6B1.07	SPAC4F8.12e	0.9999981	Known	YES	NO	NO	YES
SPAC23G3.01	SPBC28F2.12	0.9999981	Known	YES	YES	YES	YES
SPAC22F3.13	SPBC30D10.10e	0.999998	Known	YES	NO	YES	YES
SPAC458.03	SPBC216.05	0.999998	Known	NO	YES	NO	YES
SPCC23B6.03e	SPAC1687.05	0.999998	Known	NO	NO	NO	YES
SPAC15A10.16	SPCC895.05	0.999998	Known	YES	YES	YES	YES
SPAC458.03	SPBC30D10.10e	0.999998	Known	YES	NO	YES	YES
SPBC216.07e	SPAC458.03	0.999998	Known	YES	YES	YES	YES
SPAC9G1.02	SPAC1006.09	0.999998	Known	NO	YES	YES	YES
SPBC6B1.07	SPAC9.03e	0.9999979	Known	YES	NO	NO	YES
SPAC4A8.05e	SPBC1709.01	0.9999979	Known	NO	YES	YES	YES
SPAPB1E7.02e	SPBC216.05	0.9999979	Known	YES	YES	NO	YES
SPCC18B5.03	SPBP19A11.04e	0.9999979	Known	NO	YES	NO	YES
SPBC25H2.13e	SPBC16A3.11	0.9999979	Known	NO	YES	NO	NO

SPAC1783.05	SPAC3G6.01	0.9999978	Known	YES	YES	YES	NO
SPAC110.02	SPAC31A2.05c	0.9999978	Known	YES	YES	YES	YES
SPAC3C7.12	SPAC1093.06c	0.9999978	Known	YES	NO	YES	YES
SPBC19C7.03	SPBC21C3.20c	0.9999978	Known	NO	NO	YES	YES
SPBP4H10.06c	SPAC31A2.05c	0.9999978	Known	NO	NO	NO	NO
SPCC23B6.03c	SPAC16A10.07c	0.9999978	Known	NO	YES	YES	YES
SPBC342.05	SPBC216.05	0.9999978	Known	NO	YES	NO	YES
SPBC19G7.05c	SPBC21.06c	0.9999978	Known	NO	NO	NO	YES
SPBC19G7.05c	SPCC1322.12c	0.9999977	Known	NO	NO	NO	YES
SPBC25D12.03c	SPBC216.05	0.9999977	Known	NO	YES	NO	YES
SPBC19G7.05c	SPCC1739.11c	0.9999977	Known	YES	NO	NO	YES
SPAC4F8.12c	SPAC644.12	0.9999977	Known	YES	NO	NO	YES
SPAC17H9.20	SPCC5E4.04	0.9999977	Known	NO	NO	NO	NO
SPAC23E2.02	SPAC1783.05	0.9999976	Known	YES	YES	YES	NO
SPBP19A11.06	SPAC3H1.12c	0.9999976	Known	YES	YES	YES	YES
SPCC23B6.03c	SPAC9E9.08	0.9999976	Known	NO	NO	YES	YES
SPCC23B6.03c	SPBC6B1.09c	0.9999976	Known	YES	YES	YES	NO
SPBC342.05	SPBC25H2.13c	0.9999976	Known	NO	NO	YES	YES
SPAC20G8.05c	SPAC1F5.04c	0.9999976	Known	YES	YES	YES	NO
SPBC1A4.03c	SPCC5E4.06	0.9999976	Known	NO	YES	NO	YES
SPCC23B6.03c	SPAC14C4.13	0.9999976	Known	NO	YES	YES	YES
SPCC23B6.03c	SPCC126.02c	0.9999976	Known	YES	YES	YES	YES
SPCC1223.06	SPCC895.05	0.9999975	Known	YES	YES	YES	YES
SPBC216.05	SPBC29A10.15	0.9999975	Known	NO	YES	NO	YES
SPBC336.12c	SPBP19A11.04c	0.9999975	Known	YES	YES	NO	YES
SPBP4H10.06c	SPAC3H5.06c	0.9999975	Known	NO	NO	NO	NO
SPAC9.03c	SPAC644.12	0.9999975	Known	YES	NO	NO	YES
SPBC211.04c	SPCC5E4.04	0.9999975	Known	NO	NO	NO	YES
SPCC645.05c	SPBC21.06c	0.9999975	Known	NO	YES	NO	YES
SPBC19G7.05c	SPAC20G8.05c	0.9999974	Known	YES	NO	NO	YES
SPCC645.05c	SPCC1739.11c	0.9999974	Known	NO	YES	NO	YES
SPCC1223.06	SPAC15A10.16	0.9999974	Known	NO	YES	YES	YES
SPAC16A10.07c	SPBC216.05	0.9999974	Known	NO	YES	YES	YES
SPBP4H10.06c	SPBC146.03c	0.9999973	Known	YES	YES	YES	YES
SPAPYUG7.02c	SPBC30D10.10c	0.9999973	Known	YES	NO	YES	NO
SPAC13C5.07	SPBC216.05	0.9999973	Known	NO	YES	YES	YES
SPBC216.05	SPAC23C4.18c	0.9999973	Known	NO	YES	NO	YES
SPBC776.13	SPBC146.03c	0.9999973	Known	YES	NO	YES	YES
SPBC3F6.05	SPCC645.07	0.9999973	Known	YES	NO	YES	YES
SPBC31F10.14c	SPBC31F10.13c	0.9999973	Known	NO	YES	YES	YES
SPBC336.04	SPBC16D10.04c	0.9999973	Known	YES	YES	YES	YES
SPAC17D4.02	SPBC216.05	0.9999973	Known	NO	YES	NO	YES
SPBC19G7.05c	SPCC18B5.03	0.9999973	Known	NO	NO	NO	YES
SPAC1565.08	SPCC5E4.04	0.9999972	Known	YES	YES	NO	YES
SPBC29A10.04	SPAC10F6.09c	0.9999972	Known	YES	YES	YES	YES
SPBC16G5.12c	SPBC216.05	0.9999972	Known	NO	YES	NO	YES
SPBC3E7.08c	SPAC2G11.12	0.9999972	Known	NO	YES	NO	YES

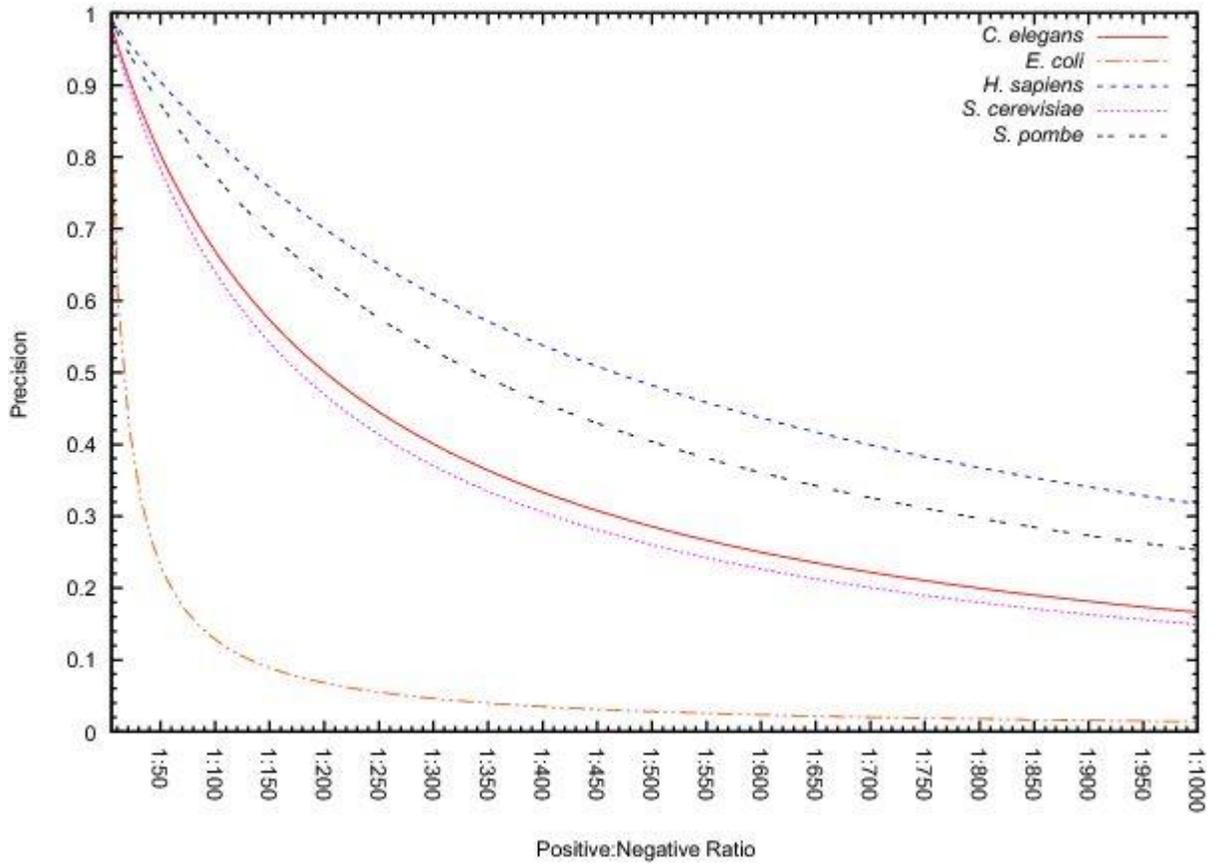
Continued

Appendix 2.2 List of interactions between 29 human proteins with established roles in efficiency of double stranded break repair and human proteome.

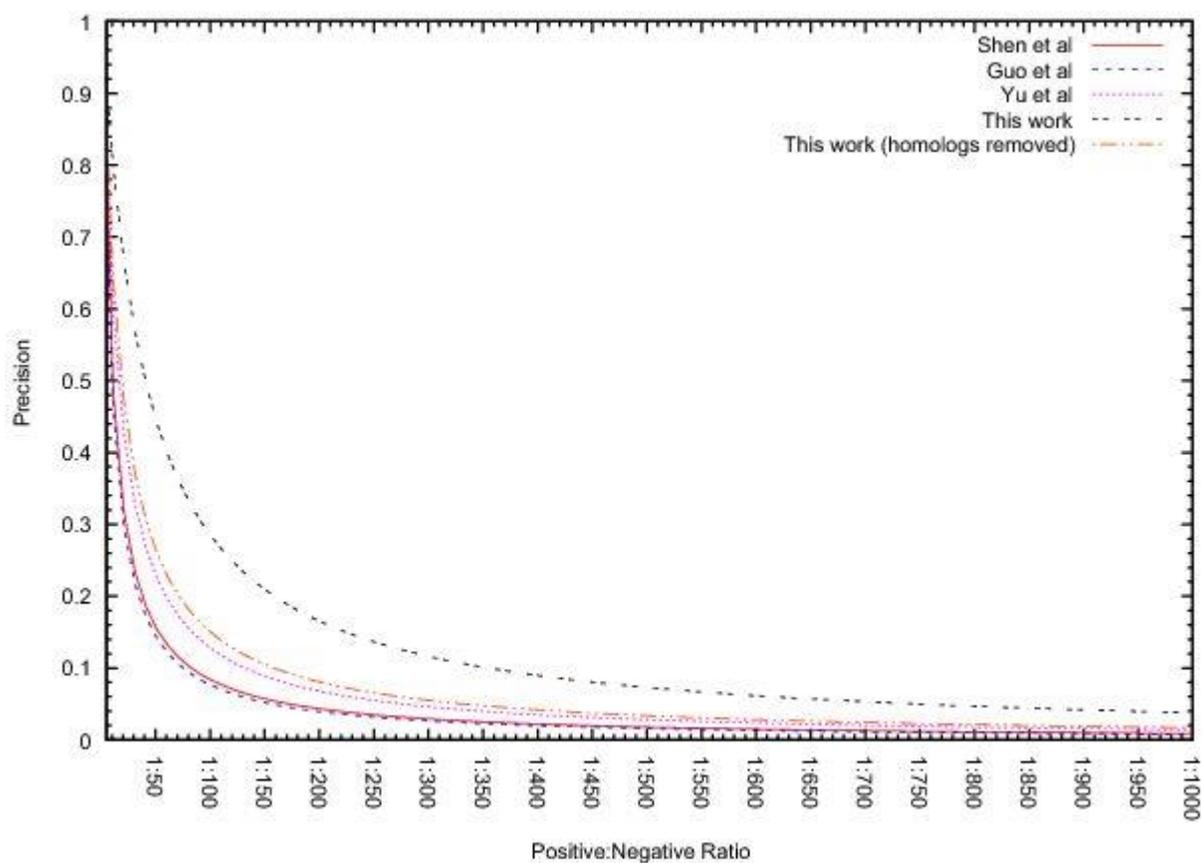
Protein A	Protein B	Score	Status
Q13315	P78527	0.99999968	Known
Q13315	P51587	0.99999961	Known
P51587	P21333	0.99999955	Known
Q13535	P51587	0.99999955	Known
Q13535	Q13315	0.9999995	Known
P78527	P38398	0.99999947	Known
Q9Y4A5	P38398	0.99999944	Known
P51587	P38398	0.99999936	Known
Q14676	Q13315	0.99999936	Known
Q13315	Q12888	0.99999933	Known
Q13315	P38398	0.99999929	Known
Q92793	Q14676	0.9999992	Known
Q14839	Q13535	0.9999992	Known
Q9BXW9	P51587	0.99999919	Known
Q14676	Q09472	0.99999919	Known
Q13535	P38398	0.99999917	Known
Q92547	Q13315	0.99999912	Known
Q92793	P38398	0.9999991	Known
Q7Z589	P51587	0.9999991	Known
Q9BXW9	Q13315	0.99999909	Known
Q09472	P38398	0.99999909	Known
Q14191	Q13315	0.99999907	Known
Q13085	P38398	0.99999907	Known
Q13315	P54132	0.99999906	Known
Q92547	O95071	0.99999904	Known
Q13315	P52701	0.99999902	Known
Q14676	P24928	0.99999901	Known
Q14676	Q12888	0.99999901	Known
Q86YC2	P51587	0.99999899	Known
Q9UHV7	P38398	0.99999899	Known
Q92878	Q13315	0.99999898	Known
Q14676	P38398	0.99999895	Known
Q14191	Q13535	0.99999892	Known
Q14683	Q13315	0.99999892	Known
P38398	P24928	0.99999889	Known
Q12888	P38398	0.99999889	Known
P51587	O60566	0.99999886	Known
Q13535	P35968	0.99999886	Known
Q9HAW4	Q13535	0.99999885	Known
Q13535	P17948	0.99999884	Known
Q13315	P35251	0.99999883	Known
Q92547	Q07864	0.99999883	Known
Q13315	P00519	0.99999881	Known
P38398	O60333	0.99999879	Known

Q13367	Q13315	0.99999876	Known
Q5SRE5	P38398	0.99999875	Known
Q92878	Q09472	0.99999871	Known
Q96PY6	P46100	0.99999869	Known
P78527	O60934	0.99999868	Known
P38398	A7E2E1	0.99999867	Novel
P51532	P38398	0.99999867	Known
P38398	B1A8Z7	0.99999864	Novel
Q13535	P00519	0.99999863	Known
Q14676	Q14191	0.99999863	Known
Q9GZU2	P38398	0.99999862	Known
P51531	P38398	0.99999862	Known
P54259	P46100	0.99999861	Known
Q15648	P38398	0.99999861	Known
Q13535	P28340	0.9999986	Known
P78527	P49959	0.99999859	Known
Q13315	Q10567	0.99999858	Known
Q13315	O94973	0.99999857	Known
P38398	P11388	0.99999857	Known
Q13315	P63010	0.99999857	Known
Q13315	P43246	0.99999856	Known
Q92831	P51587	0.99999855	Known
Q13535	P42336	0.99999855	Known
P78527	O75943	0.99999853	Known
Q9P2H0	P46100	0.99999853	Known
Q13315	P49917	0.99999852	Known
Q92878	Q14676	0.99999851	Known
Q9BXW9	P38398	0.99999851	Known
P49790	P38398	0.99999851	Known
Q13315	P49736	0.99999851	Known
Q15596	P38398	0.9999985	Known
Q99708	Q13315	0.9999985	Known
P38398	O15360	0.99999849	Known
P56524	P46100	0.99999848	Known
P51587	P42229	0.99999848	Known
Q9Y6Q9	P38398	0.99999846	Known
Q99728	P51587	0.99999845	Known
P54132	P38398	0.99999845	Known
Q14683	Q14676	0.99999841	Known
Q9ULK4	P38398	0.99999839	Known
P52701	P38398	0.99999838	Known
Q13535	Q10567	0.99999836	Known
Q9HAW4	P38398	0.99999836	Known
Q13535	O94973	0.99999834	Known
Q92547	Q02880	0.99999834	Known
Q13535	P43246	0.99999833	Known
Q92878	P38398	0.99999832	Known

Continued

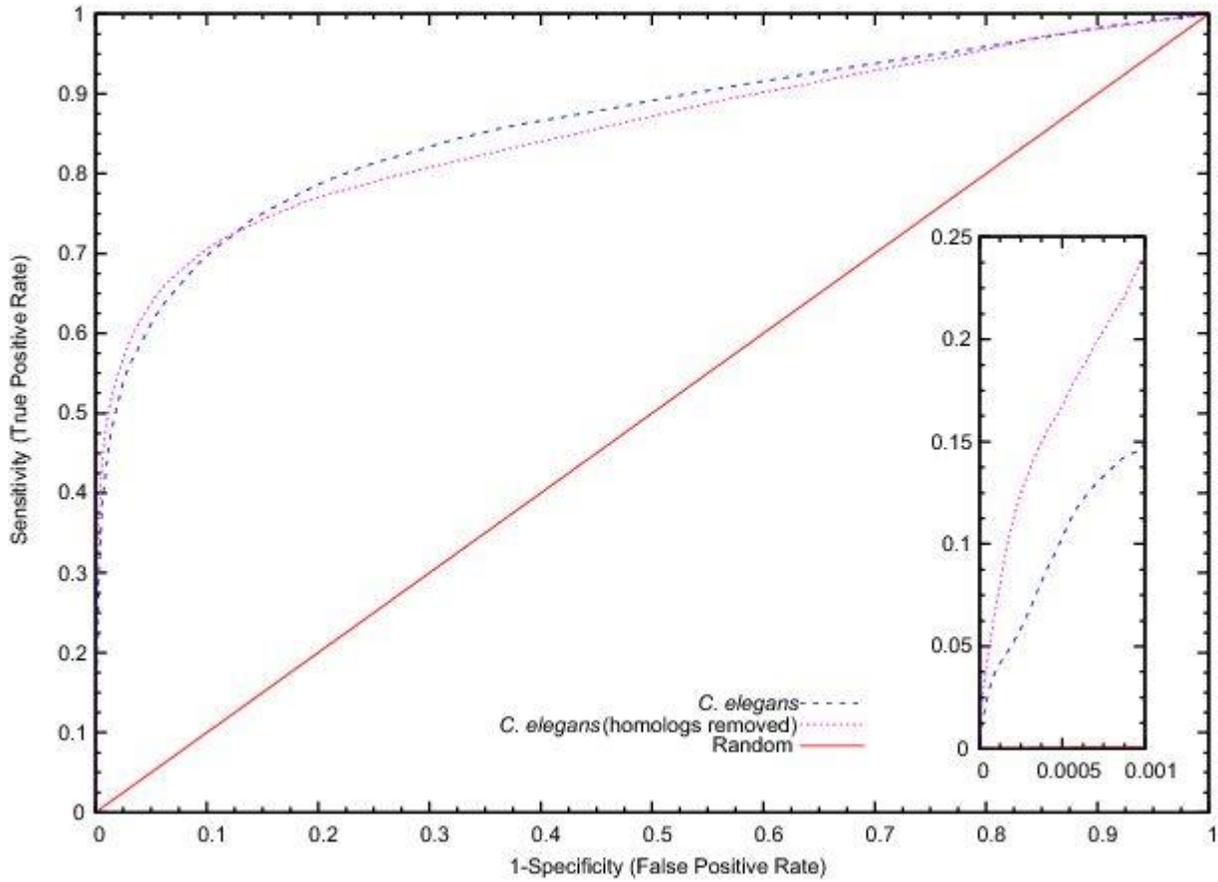


Appendix 2.3 Impact of the positive: negative ratio on the precision. Precision is reduced as positive: negative ratio is reduced.

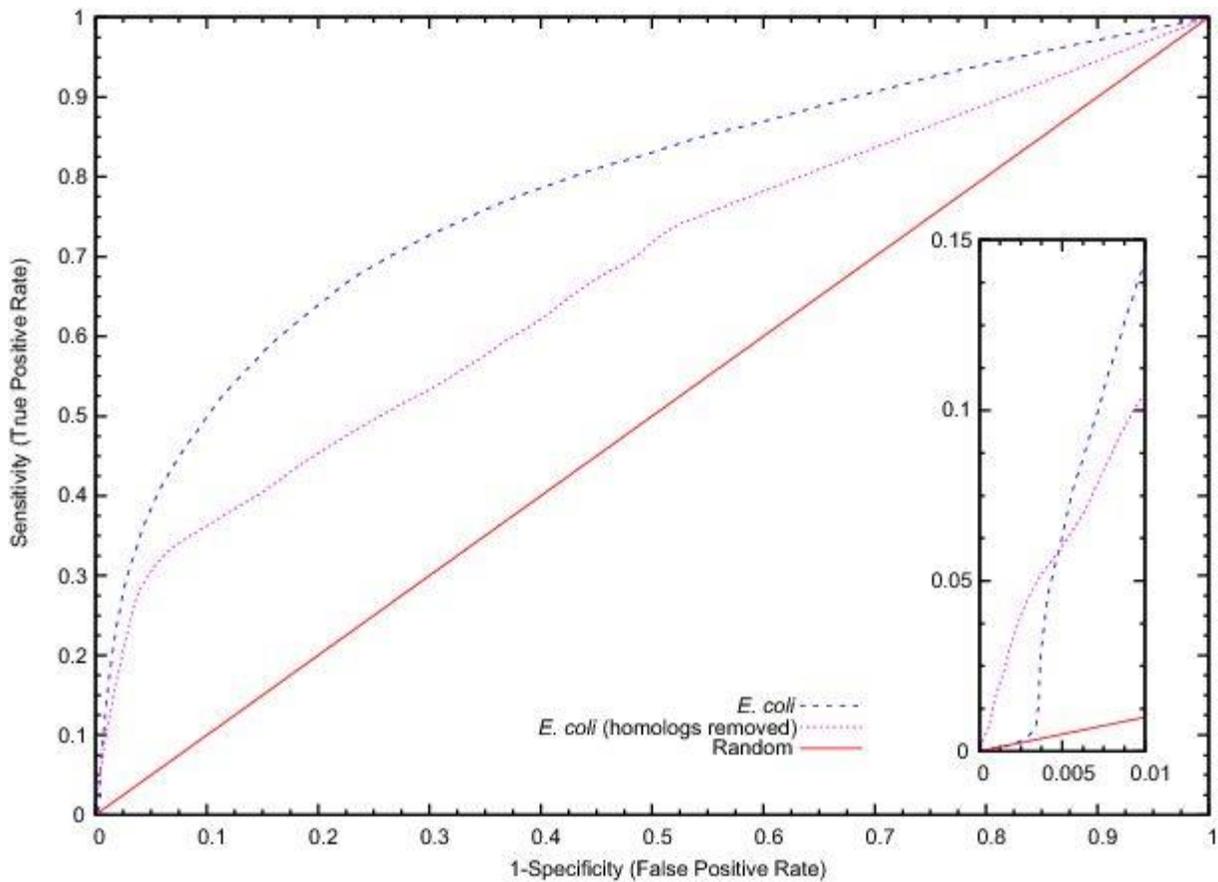


Appendix 2.4 Comparison of precision over human proteins for competing methods.

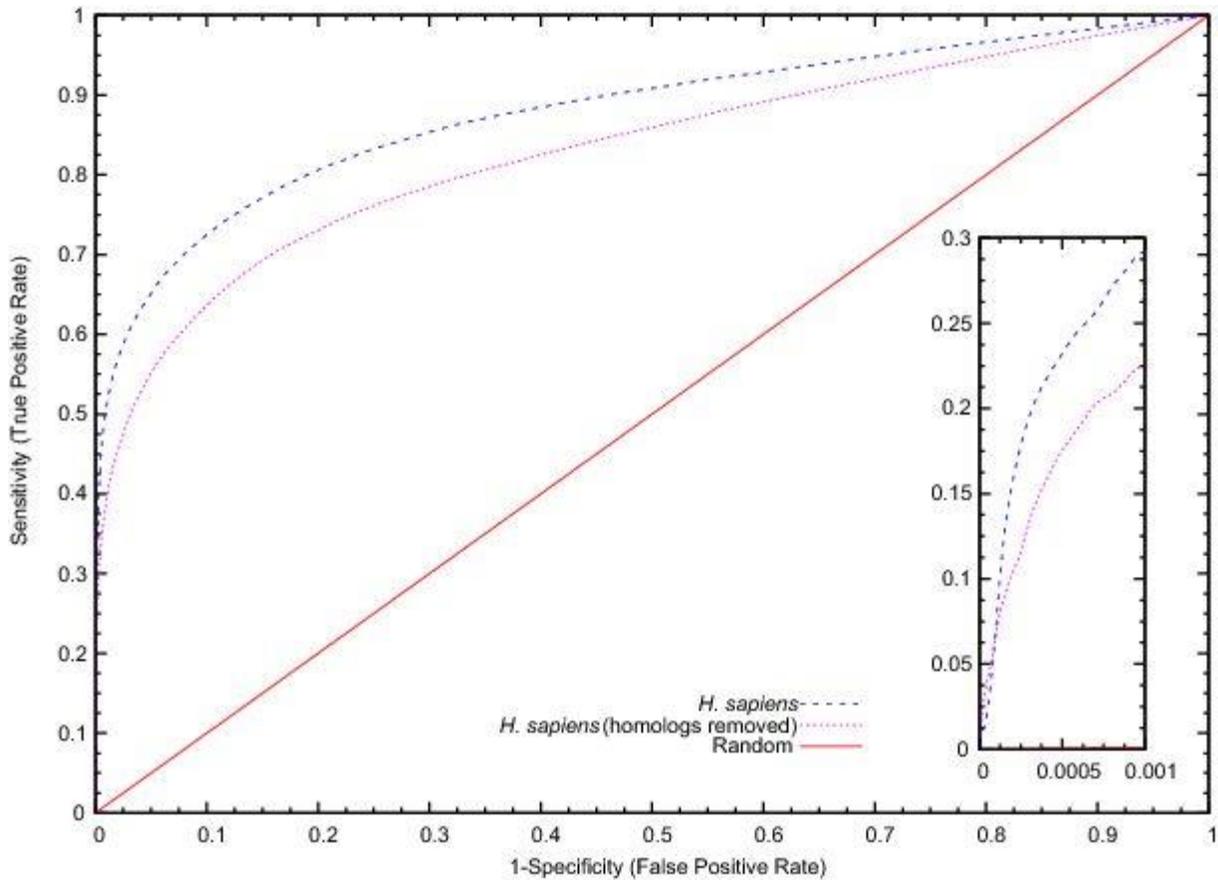
Precision for the current method (before and after the removal of homologs) is higher than those of competing methods for different positive: negative ratios. It should be noted that homologs are not removed in the competing methods.



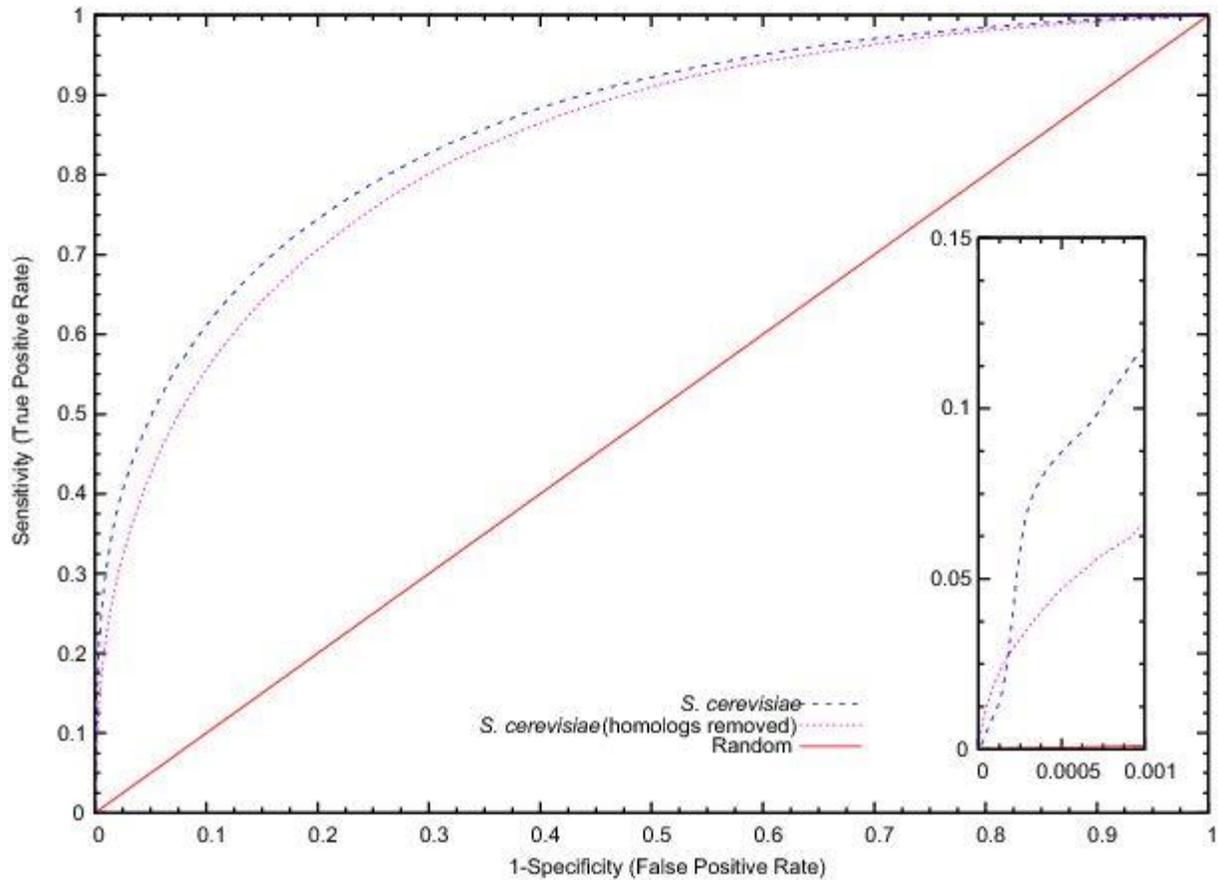
Appendix 2.5 ROC curve with and without homologous sequences for *C. elegans*. The curve presents the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). Inset: performance at very high specificity (99.95%).



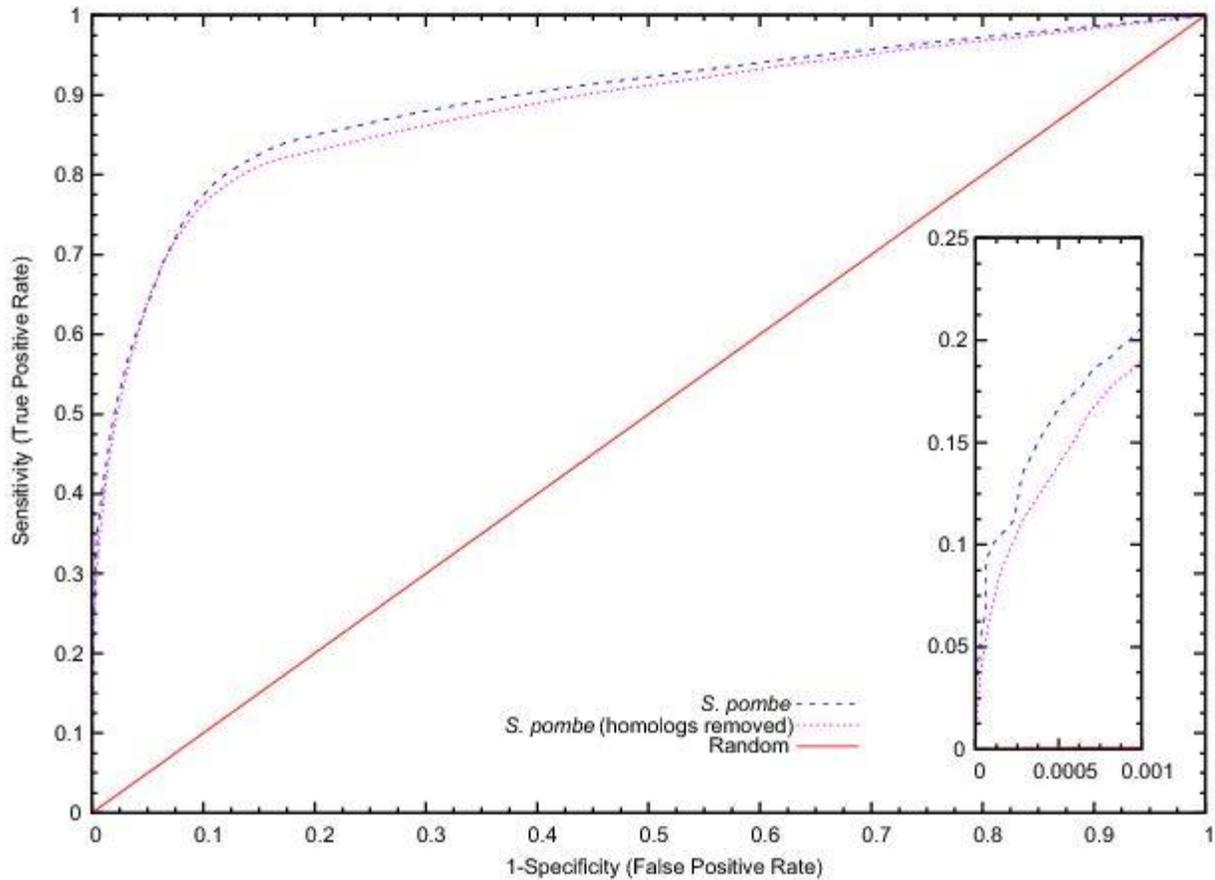
Appendix 2.6 ROC curve with and without homologous sequences for *E. coli*. The curve presents the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). Inset: performance at very high specificity (99.95%).



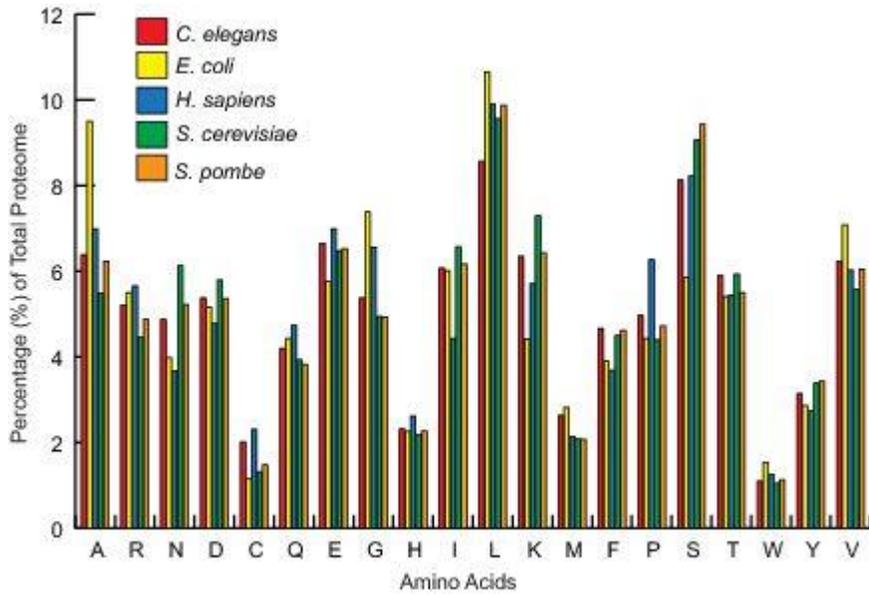
Appendix 2.7 ROC curve with and without homologous sequences for human. The curve presents the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). Inset: performance at very high specificity (99.95%).



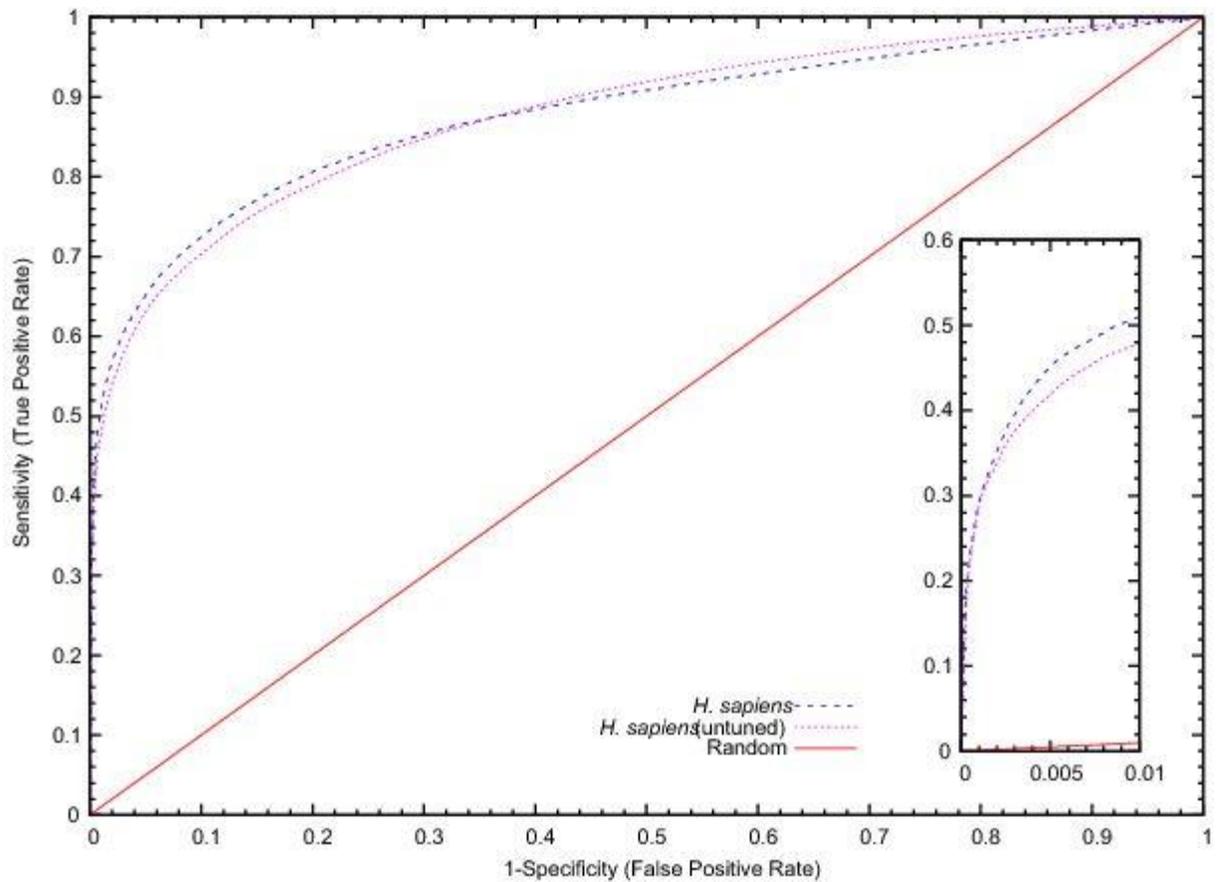
Appendix 2.8 ROC curve with and without homologous sequences for *S. cerevisiae*. The curve presents the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). Inset: performance at very high specificity (99.95%).



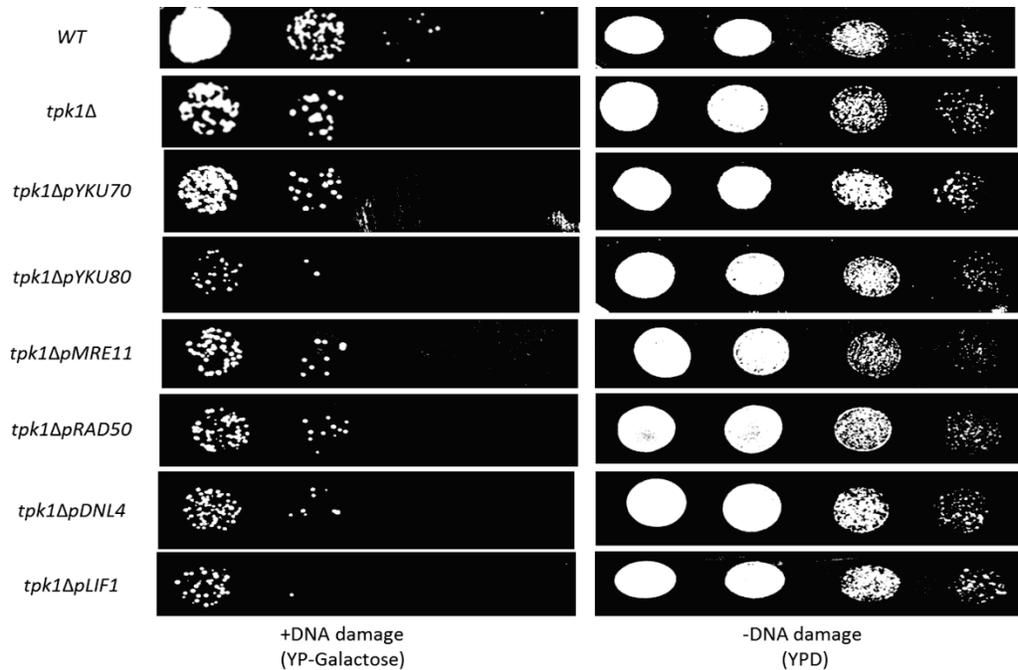
Appendix 2.9 ROC curve with and without homologous sequences for *S. pombe*. The curve presents the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). Inset: performance at very high specificity (99.95%).



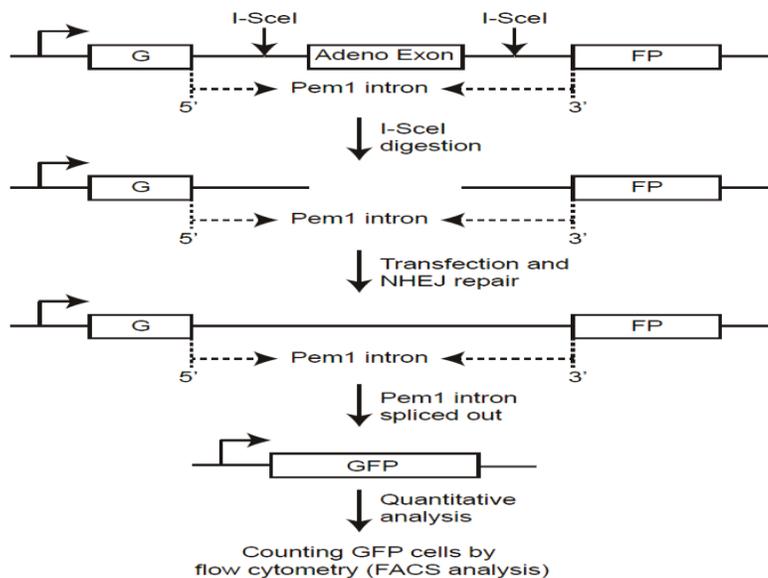
Appendix 2.10 Amino acid distributions within proteins in different organisms. The method presented in this work is tuned for these differences.



Appendix 2.11 Difference between using tuned and untuned parameters to predict human PPIs. ROC curve presents the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). Inset: performance at very high specificity (99.95%).



Appendix 3.1: Overexpression of key genes involved in NHEJ did not compensate for deletion of *TPK1* under chromosomal break induction



Appendix 3.2: Schematic representation of the pEGFP-Pem1-Ad2 reporter system used for assaying NHEJ repair efficiency (figure is modified from Wang et al, 2013). The pEGFP-Pem1-Ad2 plasmid with GFP gene interrupted by rat Pem1 intron is digested with I-SceI and transfected into U2OS cells along with the empty LentiCRISPR vector or *PRKACB* CRISPR knockout. The pre-digested pEGFP-Pem1-Ad2 in U2OS cells after transfection undergoes DSB repair. When NHEJ pathway is utilized, Pem1 intron will be sliced out, resulting in the restoration of GFP expression. Flow cytometry is then employed to calculate the number positive GFP cells in the empty vector or *PRKACB* CRISPR knockout to assess the NHEJ repair efficiency.

