EXPLORING FUNCTIONAL LANDSCAPES OF THE CELL THROUGH THE LENS OF PROTEIN INTERACTOME NETWORKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial FulFillment of the Requirements for the Degree of

Doctor of Philosophy

By

Tommy Van Vo

May 2016

EXPLORING FUNCTIONAL LANDSCAPES OF THE CELL THROUGH THE LENS OF PROTEIN INTERACTOME NETWORKS

Tommy Van Vo, Ph.D.

Cornell University 2016

Proteins function primarily by physically interacting with other proteins. As such, maps of these interactions, called protein interactome networks, can be valuable resources to help us better understand fundamental biological processes. I generated FissionNet, the first proteome-wide interactome network of the model fission yeast, *Schizosaccharomyces pombe*. FissionNet comprises 2,278 high-quality interactions, of which >90% were previously unreported in *S. pombe* and ~50% were previously not reported in any species. Moreover, FissionNet unravels previously unreported interactions implicated in processes such as stress response, gene silencing, and pre-mRNA splicing. From the evolutionary perspective, by comparing FissionNet with the interactomes of the budding yeast, *Saccharomyces cerevisiae* and of human, my colleagues Jishnu Das and Michael Meyer find that interactions are preserved better between conserved protein pairs of *S. pombe* and human compared to *S. cerevisiae* and human. To further dissect the preservation of conserved interactions, I performed large-scale cross-species interactome mapping to demonstrate that coevolution of interacting proteins is remarkably prevalent, a result with important implications for studying human disease in model organisms. Both Chapters 2 and 3 detail work done with regards to the *S. pombe* protein interactome network.

BIOGRAPHICAL SKETCH

Tommy Van Vo, the eldest of three brothers, was born and raised in Brooklyn, New York, USA to immigrant parents from the former South Vietnam. Since childhood, his parents taught him to be willing to learn anything and everything because you'll never know what will help later on in life. Growing up in the "melting pot of America", he developed a very early abstract interest in understanding how very different things can dynamically harmonize to produce outcomes greater than the sum of the individuals. This led him to pursue duo interests beginning in high school and into college at SUNY Binghamton University, economics and molecular biology. To him, the ultimate goal of both fields were nearly identical, to understand the complex relationships between individuals that give rise to incredible phenomena, whether it's societies or life itself. Ultimately, he chose to focus all his energies on following the biology path by his junior college year. Frustration with the lack of controls in economic studies compared to the relative ease of designing well-controlled biological studies made all the difference. In 2010, he joined the field of Biochemistry, Molecular and Cellular Biology (BMCB) at Cornell University with the intent of studying cancer. Almost as if that wasn't meant to be, Dr. Haiyuan Yu joined the Cornell faculty at about the same time in 2010. After a seminar by Dr. Yu who introduced Tommy to systems biology for the first time, Tommy joined the Yu lab to characterize the many molecular relationships that make up a cell and to figure out why they are important. In 2013, Tommy married his other half, Sachi Horibata, another Cornell graduate student in the field of Biological and Biomedical Sciences (BBS) who studies, ironically, the topic that systems biology lured Tommy away from: cancer.

Dedicated to my family

ACKNOWLEDGEMENTS

First, I would like to give thanks to my parents for their endless support. While they might still not fully understand what I do, they have always been there to support my dreams and ambitions. My parents fled a war-torn, oppressive regime in the late 1970s with no money, barely a high school education, and a strong willingness to start over in a country where they neither knew anyone nor the language or customs. Today, they have three sons who are all English-speakers, Americans, and college-educated. Without their unimaginable sacrifices, I would never have made it anywhere close to where I am now. The least I can do to honor them is to maintain perseverance, chase big dreams, and try to ensure that the next generation is better off than the one before. I would also like to thank my two younger brothers, my two brothers-in-law, and my parents-in-law for their undying support throughout my journey to the PhD. Thank you so much.

I would like to thank my advisor, Haiyuan Yu, for his mentorship over the past five and a half years. One of the sharpest and most energetic persons I know, he has been instrumental in terms of giving very helpful advice and challenging me to perform my utmost best. I would also like to thank my Committee members, Dave Lin and David Shalloway. They have both played major roles in shaping my approach to science and life. Dave Lin was a great collaborator and cordial mentor who gave solid support over the years for scientific, career and life issues. I have always admired and respected David Shalloway's inquisitiveness and foresight; he helped me to think outside the box and in the shape of an inverted pyramid. This way of thinking significantly affected how I did and thought about science. All three individuals had quite different styles of

mentorship but I believe this gave the perfect balance for me to succeed in graduate school and beyond. For that, I will forever be grateful.

I would like to thank all of my friends and colleagues who have provided much advice, support, interesting discussions, and kept me from being a total hermit throughout these grueling years. Some have also contributed to the work I outline in this thesis. Others have contributed to insightful discussions. Most notable are Lihua Wang, Patrice Ohouo, Francisco Bastos, Jose Ranato, Jeffrey Pleiss, Jishnu Das, and Michael Meyer. Lihua was the first person who showed me how to do experiments at Cornell and was very patient with me. Patrice, Jose, Francisco and Jeff provided much needed help with yeast biology and experiments, which were absolutely pivotal to my thesis work. Jishnu, my computational half in the Yu lab, has been trekking the PhD-route in parallel with me. Through our partnership, we have learned so much throughout these five years. He, and Michael to a lesser extent, contributed to the computational aspects of my thesis projects and helped teach me how to program. In fact, I discovered Python from Michael. There were many others who, due to lack of space, I could not explicitly mention here but I also thank them and will try to pay forward the generosity they have given me.

Finally, I would like to thank my loving wife, Sachi Horibata. She has been with me through the best of times, the worst of times, and everything in-between. From making sure I was properly fed daily, to being my discussion partner in scientific matters, to bringing reason and insight into my various trying times, she has been the one I depended on most. I truly do not believe I would have made it through the graduate school experience in one-piece without her continuous support. Thank you for believing in me and giving me the will to keep going.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

Appendix I

Appendix II

Appendix III

LIST OF TABLES

CHAPTER 2

CHAPTER 3

Appendix I

Appendix II

LIST OF ABBREVIATIONS

Term	Abbreviation
Affinity-purification followed by mass spectrometry	AP/MS
Biological process	BP
Chromatin immunoprecipitation	ChIP
Clusters of conserved eukaryotic orthologous groups of genes	KOGs
Direct coupling analysis	DCA
Double-stranded DNA	dsDNA
Gal4 activating domain	AD
Gal4 DNA-binding domain	DB
Gene Ontology	GO
High quality	HQ
Human	*H. sapiens*; H.s
Kolmogorov-Smirnov	KS
Literature-curated	LC
Mitogen activated protein kinase	MAPK
Multiple sequence alignment	MSA
Negative or random reference set	NRS; RRS
Open reading frame	ORF
Open reading frame DNA library	ORFeome
Optical density	OD
PCR-mediated linkage of adapter barcodes to nucleic acid elements with sequencing	PLATE-seq

Term	**Abbreviation**
Pearson correlation coefficient	PCC
Polymerase chain reaction	PCR
Positive reference set	PRS
Protein complementation assay	PCA
Protein-protein interaction	PPI
Reverse-transcriptase polymerase chain reaction	RT-PCR
RNA-directed RNA-polymerase Complex	RDRC
RNA-induced transcriptional silencing	RITS
Saccharomyces cerevisiae	*S. cerevisiae*; S.c
Sample size (number of biological replicates)	*n*
Schizosaccharomyces pombe	*S. pombe*; S.p
Semi-quantitative real-time polymerase chain reaction	Semi-RT-PCR
Single-stranded DNA	ssDNA
Small-scale gene duplication	SSD
Standard error	SE
Whole-genome duplication	WGD
Wild-type	WT
Wwell-based Nucleic acid programmable protein array	wNAPPA
Yeast two-hybrid	HT-Y2H; Y2H

CHAPTER 1

INTRODUCTION

1.1 SYSTEMS BIOLOGY: A NEW PARADIGM

Molecular and biochemical biology can be viewed as the study of life; life as derived from superbly complex interactions between non-living objects. Put another way, it is the economics of non-living "societies". This can arise in the form of deoxyribonucleic acids (DNA) replicating themselves by way of interacting with DNA polymerase subunits, helicases, topoisomerases, etc. Or, by way of a transcription factor binding to DNA regions and to other proteins to elicit transcriptional responses to stimuli. Together, this complexity allows for systems to arise which are both highly structured but flexible.

To rigorously dissect the inner workings of the cell, the traditional approach has been to "divide-and-conquer". The approach has been to intensely study individual pathways or processes with the goal being *in vitro* recapitulation. The idea behind this is that re-creation of processes would at least elucidate what is minimally required and sufficient to operate it and, thus, that might be the framework in the living cell. While this approach has revealed absolutely invaluable insights into cellular mechanisms, they only provide randomized and localized snapshots of the cell's inner working.

The genomics era beginning the late 20^{th} century has completely changed biological research. For the first time, we gained the ability to comprehensively identify all DNA sequences. From this information, we can predict all of an organism's genes, RNAs, and proteins. Rapid technological advancements have also allowed us to perform large-scale surveys of metabolites, DNA/protein modifications, and more. Essentially, the ability to quickly and

cheaply generate huge cellular "parts lists" has allowed us to revisit the issue of function and mechanism with a brand new perspective. Specifically, we can interrogate all components of the cell in-search of functions.

At least abstractly, molecular function can be defined as interaction(s). Since molecules in absolute isolation have no function, it must be that functions are the relationships between molecules and complexes. Large-scale dissections of interactions include gene-gene (Bajic et al., 2014; Costanzo et al., 2010; Frost et al., 2012; Ryan et al., 2012), DNA-DNA (Lieberman-Aiden et al., 2009), DNA-RNA (Chu et al., 2015), protein-modification (Ficarro et al., 2002; Rigbolt et al., 2011), protein-DNA(Johnson et al., 2007; Rhee and Pugh, 2011; Vogel et al., 2007), protein-RNA (Hafner et al., 2010), and protein-protein (Das et al., 2013; Vo et al., 2016; Yu et al., 2008). While each type of interaction tells a unique story of the cell, in this thesis, I focus on interactions between one of the most versatile and important molecular components: proteins.

1.2 PROTEIN INTERACTOMES

Systematic mappings of physical protein-protein interactions (PPIs) have garnered much attention in recent years (Singh et al., 2008). These maps (or interactomes) are usually depicted as nodes (proteins) and bi-directed edges (interactions) within network graphs (Shannon et al., 2003). At the micro level, interactomes can help to functionally characterize gene products of unknown function via the principle of "guilt-by-association", where two proteins that interact probably participate in the same or similar cellular function(s) (Oliver, 2000). At the macro level, interactomes can help to understand how biological characteristics arise from cellular network properties (Barabasi and Oltvai, 2004). At the beginning of the genomic revolution, comparisons of cellular processes at the systems-level in different organisms occurred mainly through DNA

and protein sequence similarity. It is only recently that we have been able to integrate protein-protein interaction (Arabidopsis Interactome Mapping Consortium) networks to better understand functional conservation, beyond the mere sequence level (Sharan et al., 2004).

Physical PPIs can be direct or indirect, transient or stable. Direct PPIs are biophysical interactions between two proteins. Indirect interactions refer to co-complexes where many of the components do not form binary biophysical interactions (Bader and Hogue, 2002). Various experimental methods have been developed over the years to discover interactions of these natures at the whole-genome level. Currently, the most widely-used techniques for large-scale mappings of physical binding partners are affinity purification mass spectrometry (AP-MS) and yeast two-hybrid (Y2H) (Mering et al., 2002). High-throughput AP-MS approaches in yeast have been based on purification of "natural" protein complexes by tagging a fraction of the proteome followed by component identification by matrix-assisted laser desorption/ionization with time-of-flight mass spectrometer (MALDI-TOF MS) or liquid chromatography with mass spectrometry (LC MS/MS) (Gavin et al., 2001; Ho et al., 2001). This method is useful for identifying protein sub-networks but is biased towards highly abundant and stable co-complexes (Causier, 2004).

However, Y2H approaches are based on tagging proteins-of-interest with a non-functional protein subunit. Physical binary interactions between two proteins-of-interest will bring the non-functional subunits in close enough proximity so that it becomes a functional transcription factor that can express a chromosomal reporter gene (Fields and Song, 1989). The approach is relatively cost-effective and simple compared to AP-MS because only DNA is handled by the experimenter (Vidalain et al., 2004). On a high-throughput scale, this method can be used to perform unbiased interrogations of pair-wise protein-protein interactions for entire

proteomes. This is a significant strength of Y2H because it has been shown that literature-curated interactions, which have traditionally served as a quality-control reference for high-throughput studies, can be error-prone and suffer from sociological biases (Cusick et al., 2009). In terms of cellular localization, Y2H performs quite poorly for membrane proteins but this weakness is also shared by many other protein-interaction assays (Mering et al., 2002). However, Y2H is quite successful at capturing interactions involving nuclear and cytoplasmic proteins. Additionally, Y2H can detect transient interactions even among naturally low expressed proteins (Das et al., 2012; Yu et al., 2008). Based on mRNA abundance data, Mering *et al.* found that Y2H exhibited no bias in terms of interaction coverage as a function of mRNA abundance in the studied organism (Mering et al., 2002). This is expected because mRNA levels, and probably protein levels, are controlled during Y2H by exogenous plasmid expression. Finally, work in my lab and other groups have shown that biophysical interactions obtained from Y2H are reliable because they can be validated by orthogonal, independent assays (Braun et al., 2008; Das et al., 2013; Vo et al., 2016; Yu et al., 2008). Together, these attributes of Y2H have made it the optimal approach for understanding the landscapes of protein interactomes at the resolution of binary protein-protein interactions.

Uncovering evolutionary mechanisms that shape protein interactomes is important for understanding the development of organisms and may suggest novel predictive models of how organisms can further evolve (Veron et al., 2006). Large-scale Y2H studies have been used to map PPIs for various model organisms including *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Sacchromyces cerevisiae* (Giot et al., 2003; Li et al., 2004; Yu et al., 2008). The human protein interactome has been a work-in-progress but, so far, ~50% of the PPI sample space has been systematically screened (Rolland et al., 2014; Rual et al., 2005; Yu et al., 2011).

1.3 *SCHIZOSACCHAROMYCES POMBE* AS A YEAST MODEL ORGANISM

The fission yeast is a unicellular fungus and is also a popular model organism to use due to its genetic tractability. The complete genomic sequence of the fission yeast was published in 2002 and is predicted to contain 4974 protein-encoding genes as compared to 5822 in *S. cerevisiae*, making fission yeast the eukaryote with the fewest known genes (Mewes et al., 1997; Wood et al., 2002; Wood et al., 2012). Nevertheless, both yeast species are similar in terms of sequence orthology with about 80% of *S. pombe* genes having at least one ortholog in *S. cerevisiae* (Wood et al., 2002). Based on gene sequence similarity, we might expect high levels of PPI network conservation between the species.

But, several reasons make this an unfavorable hypothesis. *S. pombe* and *S. cerevisiae* are quite distant as they are believed to have diverged ~500 to 1,000 million years ago (Balasubramanian et al., 2004; Sipiczki, 2000). In addition, despite the high degree of sequence orthology, the fission yeast is believed to be more similar to metazoans than to budding yeast (Roguev et al., 2008). Features which exemplify this are its large complex centromeric structures (35 to 110 kilobases on the three *S. pombe* chromosomes), division by fission as opposed to budding, gene and transposon regulation by the RNAi pathway, and the presence of introns in many genes (<50% of *S. pombe* genes). *S. pombe* has been significantly understudied with very little functional genomic information available. Currently, there are only 160 high quality binary PPIs known for *S. pombe* compared to 11,936 for *S. cerevisiae* (Das and Yu, 2012). Finally, Sharan *et al.* have analyzed three-way comparisons of PPI networks for *C. elegans*, *D. melanogaster*, and *S. cerevisiae* and have suggested that sequence similarity alone may not be enough to comprehensively uncover evolutionary trends of cellular circuits (Sharan et al., 2004). Altogether, these reasons show that much can be gleaned from an experimentally verified

genome-wide network of PPIs for *S. pombe*. This work will immediately benefit the *S.pombe* community and advance our understanding of fundamental principles which govern biological processes in yeast and higher eukaryotes, including human. In this thesis (Chapters 2 and 3), I will journey into the generation of the first proteome-wide fission yeast protein interactome beginning with StressNet, then expanding into the entire FissionNet. I will use the interactome (StressNet is a subset of FissionNet) to illuminate various novel protein biological functions. Additionally, through collaboration with fellow graduate students Jishnu Das and Michael Meyer, I will discuss our findings from systematic comparisons of the interactomes of *S. pombe*, *S. cerevisiae*, and human.

CHAPTER 2

CROSS-SPECIES PROTEIN INTERACTOME MAPPING REVEALS SPECIES-SPECIFIC WIRING OF STRESS RESPONSE PATHWAYS[1]

2.1 SUMMARY

The fission yeast *Schizosaccharomyces pombe* has more metazoan-like features than the budding yeast *Saccharomyces cerevisiae*, yet it has similarly facile genetics. We present a large-scale verified binary protein-protein interactome network, "StressNet," based on high-throughput yeast two-hybrid screens of interacting proteins classified as part of stress response and signal transduction pathways in *S. pombe*. We performed systematic, cross-species interactome mapping using StressNet and a protein interactome network of orthologous proteins in *S. cerevisiae*. With cross-species comparative network studies, we detected a previously unidentified component (Snr1) of the *S. pombe* mitogen-activated protein kinase Sty1 pathway. Coimmunoprecipitation experiments showed that Snr1 interacted with Sty1 and that deletion of *snr1* increased the sensitivity of *S. pombe* cells to stress. Comparison of StressNet with the interactome network of orthologous proteins in *S. cerevisiae* showed that most of the interactions among these stress response and signaling proteins are not conserved between species but are "rewired"; orthologous proteins have different binding partners in both species. In particular, transient interactions connecting proteins in different functional modules were more likely to be rewired than conserved. By directly testing interactions between proteins in one yeast species and their corresponding binding partners in the other yeast species with yeast two-hybrid assays, we found that about half of the interactions that are traditionally considered "conserved" form modified interaction interfaces that may potentially accommodate novel functions.

[1] Reprinted (with permission from the publisher) from Jishnu Das[*], Tommy V. Vo[*], Xiaomu Wei, Joseph C. Mellor, Virginia Tong, Andrew G. Degatano, Xiujuan Wang, Lihua Wang, Nicolas A. Cordero, Nathan Kruer-Zerhusen, Akihisa Matsuyama, Jeffrey A. Pleiss, Steven M. Lipkin, Minoru Yoshida, Frederick P. Roth, Haiyuan Yu. (2013). Cross-species protein interactome mapping reveals species-specific wiring of stress response pathways. ***Sci Signal*** 6(276):ra38. doi: 10.1126/scisignal.2003350. (Equal contribution is denoted by *).

2.2 CONTRIBUTIONS:

I performed Y2H ORFeome library (open reading frame collection) cloning and pairwise retests using high-throughput Y2H (HT-Y2H) (contributed to generating the *S. pombe* interactome network), PCR-stitching, stress response assays, and manuscript proofreading/editing. Jishnu Das carried out upstream experimental designs and analyses, all downstream computational analyses, and wrote most of the manuscript with Haiyuan Yu. He generated all non-experimental figures in this chapter. Xiaomu Wei performed coimmunoprecipitation experiments. Joseph C. Mellor performed Illumina sequencing. Virginia Tong performed interactome screens using HT-Y2H. Andrew G. Degatano performed interactome screens using HT-Y2H. Xiujuan Wang performed sequence alignment for Stitch-seq. Lihua Wang performed interactome screens using HT-Y2H. Nicolas A. Cordero performed interactome screens using HT-Y2H and PCR-stitching. Nathan Kruer-Zerhusen performed stress response assays. Akihisa Matsuyama generated *S. pombe* entry clones. Jeffrey A. Pleiss provided *S. pombe* strains and designed coimmunoprecipitation and stress response assays. Steven M. Lipkin designed coimmunoprecipitation and stress response assays. Minoru Yoshida directed the generation of the *S. pombe* entry clone library Frederick P. Roth co-conceived the study and oversaw Illumina sequencing. Haiyuan Yu co-conceived the study, designed all experimental and computational analyses, oversaw all aspects of the project, performed PCA and wNAPPA, and designed coimmunoprecipitation and stress response assays.

2.3 INTRODUCTION

A crucial step towards understanding properties of cellular systems is to map networks of physical DNA-, RNA-, and protein-protein interactions, or the "interactome network," of an organism. Over the last decade, large-scale binary protein-protein interactome datasets have been produced for several eukaryotes – *Saccharomyces cerevisiae* (Ito et al., 2001; Uetz et al., 2000; Yu et al., 2008), *Drosophila melanogaster* (Formstecher et al., 2005; Giot et al., 2003), *Caenorhabditis elegans* (Li et al., 2004; Simonis et al., 2009), *Arabidopsis thaliana* (Arabidopsis Interactome Mapping Consortium, 2011), and human (Rual et al., 2005; Stelzl et al., 2005), among which we produced a high-quality whole-proteome interactome network in *S. cerevisiae* using a high-throughput yeast two-hybrid (HT-Y2H) system (Yu et al., 2008). However, due to large evolutionary distances between these species (the last common ancestor of fungi and human is over 1 billion years ago (Sipiczki, 2001; Wood et al., 2002)) and extremely low coverage (where most protein interactions are yet to be detected) of available interactome maps outside of *S. cerevisiae*, the overlap between these networks is sparse (Gandhi et al., 2006). This makes it difficult to extract meaningful information about evolutionary relationships from these interactomes. Thus, to bridge this gap, it is essential to construct a high-coverage interactome network for an intermediate species. The fission yeast, *Schizosaccharomyces pombe*, has one of the most easily manipulatable genomes, and it is estimated to have diverged from the budding yeast, *S. cerevisiae*, approximately 400 million years ago (Sipiczki, 2000; Wood et al., 2002). Furthermore, fission yeast is more similar to metazoans than is budding yeast, especially in its gene regulation by chromatin modification and RNA interference, mechanisms that are absent in budding yeast (Roguev et al., 2008). A high-quality map of the protein-protein interactome

network of *S. pombe* will enable analysis of biological properties of many complex pathways common in metazoan species but missing in *S. cerevisiae* (Shevchenko et al., 2008).

Moreover, the two yeasts live in highly disparate ecological niches and have varied mechanisms of responding to external stimuli. Therefore, in this study, we focus on 658 *S. pombe* genes involved in key regulatory processes of stress response and cellular signaling. These pathways control how organisms sense and adapt to their immediate environments, and are therefore likely to have diverged between the two species. Using our HT-Y2H pipeline (Yu et al., 2008), we obtain a binary interactome network among these genes, which we named "StressNet". All interactions were verified with two orthogonal assays to ensure their quality. By comparing with their *S. cerevisiae* counterparts, we first measure the conservation rate of these StressNet interactions between fission and budding yeasts using a Bayesian method. We find significant species-specific wiring of stress-response and signaling pathways far beyond what is expected by sequence orthology, indicating that rewiring of protein interactome networks in related species is likely to be a major factor for divergence. We also identify a previously unknown component of the Sty1 mitogen activated protein kinase (MAPK) pathway and experimentally validated that it has gained novel functions through rewiring of its interactions. Furthermore, to better understand the evolution of proteins and their interactions, we developed a large-scale cross-species interactome mapping approach to directly test interations between *S. pombe* proteins and the *S. cerevisiae* orthologs of their partners. Such analysis is only possible with the availability of two well-controlled high-coverage interactome maps generated with the same technology. We find that, for many conserved interactions, both partners have co-evolved to accommodate novel interactions and functions, and their interaction interfaces can no longer be recognized by their *S. cerevisiae* counterparts.

2.4 MATERIALS AND METHODS

2.4.1 Selection of genes for the study

This study focused on stress response and signal transduction proteins (based on GO Biological Process annotations) and their known interactors in *S. pombe*. We also included *S. pombe* orthologs of *S. cerevisiae* proteins that are known to interact with orthologs of fission yeast stress response and signal transduction proteins. While selecting the 658 ORFs, we also ensured that a set of PRS interactions in *S. pombe* could be constructed with genes from our space, a limiting criterion because there are only 160 binary high-quality *S. pombe* interactions reported in the literature.

2.4.2 Yeast two-hybrid

Y2H experiments were carried out as described (Yu et al., 2011). Briefly, 658 *S. pombe* ORFs in Gateway entry vectors were transferred into AD and DB vectors using Gateway LR reactions. After bacterial transformation, plasmids of all AD-Y and DB-X clones were transformed into Y2H strains *MAT**a*** Y8800 and *MATα* Y8930 (genotype: *leu2-3, 112 trp1-901 his3D200 ura3-52 gal4D gal80D GAL2-ADE2 LYS2::GAL1-HIS3 met2::GAL7-lacZ cyh2R*), respectively. The *MAT**a*** Y8800 strain was obtained from the *MAT**a*** Y550 strain after mutating *CYH2* to introduce cycloheximide (CHX) resistance. *MATα* Y8930 was generated by crossing *MAT**a*** Y8800 with *MATα* Y1541 (Uetz et al., 2000), followed by sporulation and identification of the *MATα* CHX-resistant yeast strain by tetrad analysis. After AD-Y and DB-X were transformed into Y8800 and Y8930, respectively, autoactivators were screened by spotting onto synthetic complete medium (SC) lacking histidine and tryptophan (AD-Y) or histidine and leucine (DB-X). These autoactivators were excluded from all further screenings. Each unique

DB-X was mated with pools of ~188 unique AD-Y by co-spotting onto yeast extract peptone dextrose (YEPD) plates. Diploids were selected by replica plating onto SC plates without leucine and tryptophan (SC –Leu −Trp). To select for positive interactions, we performed Y2H screening by replica plating the diploids onto SC plates with 1 mM 3-amino-1,2,4-triazole (3-AT) and without leucine, tryptophan, and histidine (SC –Leu –Trp –His +3-AT). SC –Leu –Trp −His plates were used for the HT-Y2H screen in *S. cerevisiae* (1). We used 1 mM 3-AT because this concentration greatly reduces background and improves the quality of the screens (Arabidopsis Interactome Mapping Consortium, 2011; Venkatesan et al., 2009; Yu et al., 2011). Newly occurring autoactivators were determined by concurrently replica plating the diploids onto SC medium with CHX and 1 mM 3-AT and lacking leucine and histidine (SC –Leu –His +3-AT +CHX). Screening for these autoactivators relies on CHX to select for cells that do not have the AD plasmid because of plasmid shuffling. Thus, growth on the latter plate identifies spontaneous autoactivators; these were removed from further analyses. All plates were replica cleaned the following day and scored after three additional days. The space was screened three times.

Y2H positives were grown 2 to 3 days at 30°C and then spotted onto four plates for secondary phenotype confirmation (phenotyping II) (SC –Leu –Trp –His +3-AT; SC –Leu –His +3-AT +CHX; SC –Leu –Trp −adenine; SC –Leu –adenine +CHX). Colonies that either grew on SC –Leu –Trp –His +3-AT but not on SC –Leu –His +3-AT +CHX or grew on SC –Leu –Trp –adenine but not on SC –Leu –adenine +CHX were identified as positives.

For colonies that scored positive in phenotyping II, the identities of DB-X and AD-Y were determined by the Stitch-seq approach (Yu et al., 2011) using Illumina sequencing. All identified interacting pairs were retested by pairwise Y2H.

2.4.3 Construction of positive (PRS) and random (RRS) reference sets

The PRS and RRS are representatives of true-positive interactions and negative pairs, respectively, and we used the PRS and the RRS to optimize the assay performance, and they may be interpreted as positive and negative controls. The PRS comprises a set of 54 protein interactions from the literature, each of which is supported by at least two independent assays from two different publications. RRS pairs were generated from a random selection out of all possible protein pairs within our search space for which no interaction has yet been detected by any method. Because fission yeast interactions are underexplored, we also required that their corresponding budding yeast ortholog pairs have never been reported to interact. Another way to construct the RRS is to consider protein pairs with different cellular localizations because these are unlikely to interact. Thirty-one of the 43 RRS pairs are indeed localized in different cell compartments. Using the whole RRS (Figure 2.1C), we estimated the false-positive rates for Y2H, PCA, and wNAPPA to be 0/43, 2/43 (4.7 ± 3.2%), and 2/43 (4.7 ± 3.2%), respectively. If we only use the 31 RRS pairs localized in different cell compartments (named "RRS_DiffLocal"), the false-positive rates for the three assays are 0/31, 2/31 (6.5 ± 4.4%), and 1/31 (3.2 ± 3.2%). Therefore, the false-positive rates for all three assays used in our experiments do not change whether we use the complete RRS or RRS_DiffLocal. With these controls, we found that 20 of the 54 PRS, and none of the RRS, were detected in our screen. We calculated the sensitivity of our assay as 20/54 (37.0 ± 4.4%).

2.4.4 Protein complementation assay

S. pombe ORFs available in Gateway entry vectors were transferred by Gateway LR reactions into vectors encoding the two fragments of YFP (Venus variant) fused to the N

terminus of the tested proteins. Baits were fused to the F1 fragment (amino acids 1 to 158 of YFP), and preys were fused to the F2 fragment (amino acids 159 to 239 of YFP). After bacterial transformation, plasmid DNA was prepared on a Tecan Freedom Evo biorobot, and DNA concentrations were determined by the absorbance at 260 nm (A260) with a Tecan M1000 in a 96-well format. A 50-ng aliquot of each vector encoding the two proteins was used for transfection into human embryonic kidney 293T cells in 96-well plates, using Lipofectamine 2000 (Invitrogen) reagent according to the instructions of the manufacturer. At about 48 hours after transfection, cells were processed with a Tecan M1000. A pair is considered interacting if the YFP fluorescence intensity was ≥2-fold higher over background.

2.4.5 Well-based nucleic acid programmable protein array

ORFs encoding the interacting proteins were cloned into Gateway-compatible pCITE-HA (hemagglutinin) and pCITE-GST (glutathione S-transferase) vectors by LR reactions. After bacterial transformation, growth, DNA minipreps, and determination of DNA concentration, ~0.5 μg of each plasmid was added to Promega TnT coupled transcription-translation mix (catalog no. L4610) and incubated for 90 min at 30°C to express proteins. During this time, anti-GST antibody–coated 96-well plates (Amersham 96-well GST detection module, catalog no. 27-4592-01) were blocked at room temperature with phosphate-buffered saline containing 5% dry milk powder. After protein expression, the expression mix was diluted in 100 μl of blocking solution and added to the emptied pre-blocked 96-well plates. Expression mix was incubated in the 96-well plates for 2 hours at 15°C with agitation to allow for protein capture. After capture, plates were washed three times and developed by incubation with primary and secondary antibodies. Signal was visualized by chemiluminescence (Amersham ECL reagents, catalog

no.RPN2106) with a Tecan M1000 plate reader. Wells with ≥3-fold higher intensity over background in either configuration were considered positives.

2.4.6 Measuring the precision of our assay

The precision of the Y2H assay was measured experimentally using either PCA or wNAPPA. The precision of the Y2H assay was calculated using PCA and wNAPPA as orthogonal validation assays. Using Bayes' rule we can build relationships between true and false positive rates of Y2H and observed positive interactions by a validating assay as:

$$\Pr(A+|Y+) = \Pr(A+|Y+,T+) \times \Pr(T+|Y+) + \Pr(A+|Y+,T-) \times \Pr(T-|Y+)$$

where *A*+ corresponds to observing a positive interaction using the validating assay, *Y*+ corresponds to observing a positive interaction using Y2H, and *T*+ (*T*−) corresponds to an interaction being a real positive (negative) interaction. The precision of the Y2H is the term Pr(*T*+|*Y*+) [which is also equal to 1− Pr(*T* −|*Y* +)].

Assuming conditional independence between the validating assay and Y2H based on previously defined reasons (Yu et al., 2008), we can write:

$$\Pr(A+|Y+) = \Pr(A+|T+) \times \Pr(T+|Y+) + \Pr(A+|T-) \times \Pr(T-|Y+)$$

Solving for the precision of the Y2H assay yields:

$$\Pr(T+|Y+) = \frac{\Pr(A+|Y+) - \Pr(A+|T-)}{\Pr(A+|T+) - \Pr(A+|T-)}$$

Pr(*A*+|*T*+) and Pr(*A*+|*T*–) were measured in the PRS and RRS experiments. So, for our Y2H assay we can write precision as:

$$Precision = \frac{F_{StressNet} - F_{rrs}}{F_{prs} - F_{rrs}}$$

where $F_{StressNet}$ is the fraction positive by an assay for StressNet, which is the best estimator for Pr(T+|Y+). F_{PRS} is the fraction positive by the assay for the PRS, which is an estimator for Pr(A+|T+). F_{RRS} is the fraction positive by the assay for the RRS, which is an estimator for Pr(A+|T-).

The standard errors of $F_{StressNet}$, F_{prs}, and F_{rrs} are calculated using the standard error for binomial distributions:

$$StdErr = \sqrt{\frac{F(1-F)}{N}}$$

where F is the fraction positive by the assay ($F_{StressNet}$, F_{prs}, or F_{rrs}) and N is the total number of pairs tested.

To estimate the standard error for the precision, we used the standard delta method:

$$\sigma_X^2 = (\frac{\partial f}{\partial A}\sigma_A)^2 + (\frac{\partial f}{\partial B}\sigma_B)^2 + (\frac{\partial f}{\partial C}\sigma_C)^2 + ...$$

where $X = f(A, B, C, \ldots)$. $A, B, C, \ldots$ are independent random variables.

Here, the standard error of the precision is calculated as:

$$\sigma_{precision} = \sqrt{(\frac{1}{F_{prs} - F_{rrs}})^2 \times \sigma_{StressNet}^2 + \frac{(F_{StressNet} - F_{rrs})^2}{(F_{prs} - F_{rrs})^4} \times \sigma_{prs}^2 + \frac{(F_{StressNet} - F_{prs})^2}{(F_{prs} - F_{rrs})^4} \times \sigma_{rrs}^2}$$

We have two validating assays, and we can incorporate the precision rates from these assays by calculating the average precision:

$$Average\ Precision = \frac{Precision_{PCA} + Precision_{wNAPPA}}{2}$$

The standard error for the average precision is calculated by the delta method as:

$$\sigma_{average\ precision} = \sqrt{\frac{\sigma_{PCA}^2}{4} + \frac{\sigma_{wNAPPA}^2}{4}}$$

Using this framework, we estimate the precision of our Y2H assay to be $95.3 \pm 4.7\%$.

2.4.7 Calculating confidence scores for interactions

Using the random forest algorithm (Breiman, 2001), we integrated the results from Y2H, PCA, and wNAPPA and calculated the confidence scores for interactions. Random forest is an ensemble classifier that constructs multiple decision trees by stochastic discrimination (Kleinberg, 1996) and predicts a final class on the basis of a weighted combination of the output class of each decision tree. It is considered to be a robust and accurate classifier for noisy data sets (Breiman, 2001).We evaluated the performance of our classifier by fivefold cross-validation on our reference set (union of PRS and RRS) and obtained moderately good performance (AUC = 0.64; Figure 2.2B-D).

2.4.8 Determination of orthologs between *S. pombe* and *S. cerevisiae*

We used the list of orthologs provided by PomBase (Wood et al., 2012). The genome of *S. cerevisiae* underwent a duplication event (Kellis et al., 2004). Thus, many *S. pombe* genes

have two corresponding *S. cerevisiae* orthologous genes. Moreover, in a number of cases, the same *S. cerevisiae* gene has multiple *S. pombe* orthologs. Thus, the mapping considered for the study was "many-to-many."

2.4.9 Estimation of the conservation of interactions

To estimate the conservation rate of protein-protein interactions between *S. pombe* and *S. cerevisiae*, we used a Bayesian framework that incorporates the precision and sensitivity of our Y2H assay:

$$\Pr(Det) = \Pr(Det \mid Cons+) \times \Pr(Cons+) + \Pr(Det \mid Cons-) \times \Pr(Cons-)$$

The symbols have the same meanings as defined in the Results section.

By definition:

$$\Pr(Cons+) = 1 - \Pr(Cons-)$$

We can simplify the earlier equation to obtain an expression for Pr(*Cons*+):

$$\Pr(Cons+) = \frac{\Pr(Det) - \Pr(Det \mid Cons-)}{\Pr(Det \mid Cons+) - \Pr(Det \mid Cons-)}$$

To estimate the error for the conservation percentage, we used the standard delta method as described earlier. The standard deviation of *Cons*+ is given by:

$$\sigma_{Cons+} = \sqrt{\frac{(F_{prs} - F_{rrs})^2 \sigma_{F_{det}}{}^2 + (F_{det} - F_{rrs})^2 \sigma_{F_{prs}}{}^2 + (F_{det} - F_{prs})^2 \sigma_{F_{rrs}}{}^2)}{(F_{prs} - F_{rrs})^4}}$$

It is indeed possible to obtain a measure of the conservation rate by inverting the above setup. We mapped all *S. pombe* proteins in our space to their corresponding *S. cerevisiae* orthologs. We calculated the number of interactions in this *S. cerevisiae* space detected by our Y2H assay. We then mapped all the observed *S. cerevisiae* interactions to their corresponding *S. pombe* ortholog pairs and calculate the number of pairs detected as interacting in StressNet. We find that for 48/386 (12.4%) *S. cerevisiae* interactions, the corresponding *S. pombe* ortholog pairs also interact. Using the Bayesian framework described above, we calculate the conservation rate between *S. pombe* and *S. cerevisiae* to be 34.7 ± 2.0%.

2.4.10 Interaction conservation and confidence scores

After supplementing our Y2H experiments with high-quality interactions from the literature, we find that 90/235 (38.3%) interactions are conserved in StressNet. The statistical error associated with this measurement is related to the sample size and is calculated as the standard error [standard error = standard deviation/square root (N), where N is the number of samples]. The standard deviation is calculated on the basis of the underlying probability distribution. The conservation percentage is obtained by a simple division (90/235 = 38.3%), and the underlying probability distribution is binomial (because each interaction can either be conserved or not, it corresponds to a Bernoulli event, the ensemble of which is modeled by a binomial distribution). The standard error is calculated using the appropriate formula for a binomial distribution: square root [$p \times (1 - p)/N$] = 3.2%, where p is the fraction of interactions that are conserved (90/235) and N is sample size (235). To test whether interactions with higher confidence scores were more likely to be conserved, we divided all StressNet interactions into two groups. The first group comprises interactions with confidence scores in the lower two

quartiles, and the second group comprises interactions with confidence scores in the upper two quartiles. We then compared the conservation for these two groups. We find that there is no significant difference (P = 0.37 using a two-sided Fisher's exact test) in conservation rate between the two groups. This validates that the observed conservation rate is robust and not correlated with the confidence score associated with each interaction.

2.4.11 Evolutionary rates of genes and protein interactions

The evolutionary rate of genes is commonly measured in terms of the ratio of asynchronous nucleotide substitutions per asynchronous site to synchronous substitutions per synchronous site, or *dN/dS*. This quantifies the selective evolutionary pressure on certain protein-coding genes to diverge faster, as opposed to others that may almost remain unchanged across species (Ceol et al., 2010). To calculate the *dN/dS* values for all *S. pombe* genes, we used two sequenced species in the *Schizosaccharomyces* genus— *S. cryophilus* and *S. octosporus* (Hu et al., 2007; Mewes et al., 2011). To determine orthology relationships, we used BLAST-x with default parameters (Cusick et al., 2009) on all *S. pombe* genes. The top BLAST hit for each *S. pombe* gene against the indexed database of proteins for each of the two species was designated to be an ortholog, provided the E-value of the hit was <0.05. Although the E-value cutoff is relatively high, it ensures that no potential pairs are missed. For pairs that have been incorrectly estimated to be orthologs, there is a correction step in downstream calculations that will return a *dN/dS* value of NaN (not a number) because of too high divergence. For all orthologous pairs, the Nei-Gojobori algorithm (Ceol et al., 2010), which uses the Jukes-Cantor substitution model, was used to calculate *dN/dS* values.

2.4.12 Conservation of interactions and sequence similarity

Sequence similarity between *S. pombe* ORFs and their *S. cerevisiae* orthologs was measured by performing pairwise sequence alignment between all known ortholog pairs. This was performed using the Needle program in the EMBOSS suite (Rice et al., 2000). It uses the Needleman-Wunsch alignment algorithm (Needleman and Wunsch, 1970) to find the optimum alignment of two sequences along their entire length. The recommended default parameters – an affine gap penalty model (Vingron and Waterman, 1994) with an opening penalty of 10 and an extension penalty of 0.5 and the BLOSUM62 scoring matrix (Henikoff and Henikoff, 1992) – were used for the alignment. Since the lengths of orthologs may be dissimilar, we calculated the overall similarity percentage (Arabidopsis Interactome Mapping Consortium, 2011) with reference to the length of the *S. pombe* ORFs:

$$OPS = \frac{N_{st}}{L\_Sp_t}$$

where, N_{st} is the total number of similar residues and L_Sp_t is the total length of the *S. pombe* ORF.

We then examined the relationship between the similarity percentage and the percentage of conserved interactions. Since the number of interactions varies considerably across different bins, we require each bin to have at least 5 interactions. If any bin has less than 5 interactions, it is merged with the next (higher) bin. This ensures that our results are robust to outlier effects. We found that there is an increase in the degree of conservation with an increase in overall sequence similarity. To examine if the primary cause of this trend is the similarity of conserved domains, we identified domains on ortholog pairs that interact (Finn et al., 2005; Stein et al., 2011). We defined the percentage similarity of interacting domains (PSID) as:

$$PSID = \frac{N_{si}}{L\_Sp_i}$$

where N_{si} is the number of similar residues in interacting domains and L_Sp_i is the sum of the lengths of the interacting domains in *S. pombe*.

We find that there is no correlation between PSID and degree of conservation. We also repeated our analysis using sequence identity instead of similarity and the results remain unchanged. This suggests that conservation of interactions is more complex than previously imagined (Espadaler et al., 2005; Kim et al., 2006) – along with the actual interfaces, neighboring secondary structures may also play a major role in determining the thermodynamic feasibility of the interaction.

2.4.13 Inferring interaction interfaces from 3did and iPfam

In this study, we use interacting domains identified by 3did (Stein et al., 2011) and iPfam (Finn et al., 2005) to define interaction interface. To verify the reliability of inferring these domain-domain interactions, we performed three-fold cross-validation for 1,456 interaction pairs that have co-crystal structures. Since there are very few co-crystal structures for *S. pombe*, this approach allowed us to obtain a meaningful estimate of the quality of the domain-domain predictions in these two databases. We split the pairs into three subsets such that two subsets are used for training and the third one is the test set. For each interaction pair in the test dataset, we scored a successful structural prediction when the predicted domain-domain interaction(s) had at least one co-crystal structure in support of it. We repeated the procedure thrice with each of the three subsets as the test set. Among the 1,456 PPI pairs, over 90% can be correctly predicted with corresponding interacting domains. This analysis indicates the predicted interaction interfaces used for our calculations are indeed high quality (Wang et al., 2012).

2.4.14 Robustness of differences between sets of conserved and rewired interactions

To assess the robustness of the differences between sets of conserved and rewired interactions, we constructed different sets of conserved and rewired interactions corresponding to different confidence levels.

We constructed two sets of conserved interactions at different confidence levels – Conserved_HQ and Conserved_All. Conserved_HQ comprises only those interactions whose corresponding *S. cerevisiae* ortholog pairs have tested positive in our Y2H experiments or have been confirmed to interact by two or more independent orthogonal assays in the literature, while Conserved_All comprises all interactions in Conserved_HQ and those *S. cerevisiae* ortholog pairs that have been reported as interacting in the literature by only one assay.

We constructed five sets of rewired interactions at different confidence levels – Rewired_ByDefn, Rewired_HQ, Rewired_LC, Rewired_All_DiffLocal, and Rewired_All. Rewired_ByDefn comprises only those StressNet interactions where at least one of the interacting proteins does not have a *S. cerevisiae* ortholog and, therefore, no corresponding interaction can exist in *S. cerevisiae*. Thus, these interactions are rewired by definition. Rewired_HQ comprises all interactions in Rewired_ByDefn and those interactions whose corresponding *S. cerevisiae* ortholog pairs are known to have other high-quality interactions but have never been reported as interacting in the literature or tested positive in our Y2H experiments, and these ortholog pairs are known to have different cellular localizations. Thus, these correspond to *S. pombe* interactions whose corresponding budding yeast ortholog pairs are in principle non-interacting, as they have different cellular localizations (Jansen et al., 2003; Yu et al., 2008) and they participate in well-validated interactions with other proteins (thus, they have been previously explored to identify their interaction partners) but have never been reported

to interact in the literature. Rewired_LC comprises all interactions in Rewired_ByDefn and those interactions whose corresponding ortholog pairs are known to have other high-quality interactions but have never been reported as interacting in the literature or tested positive in our Y2H experiments. Rewired_All_DiffLocal corresponds to all interactions in Rewired_ByDefn and those interactions whose corresponding ortholog pairs have different cellular localizations, and Rewired_All comprises all interactions that are not in Conserved_All.

All our results remain the same for all sets of conserved and rewired interactions at different confidence levels, indicating that the observed differences between these two sets of interactions are robust and reliable.

2.4.15 Construction of myc-sty1 and HA-snr1 expression clones

S. pombe sty1 and *snr1* genes were PCR amplified using the following primers – sty1-pNCH1472-Forward, sty1-pNCH1472-Reverse, snr1-pSGP73-Forward, and snr1-pSGP73-Reverse (Table 2.1). The sty1 PCR product was cloned into a pNCH1472-myc vector via NotI and SalI restriction sites. The *snr1* PCR product was cloned into a pSGP73-HA vector via NotI and BglII restriction sites. pNCH1472-myc-sty1 and pSGP73-HA-snr1 were single or double transformed into *S. pombe* KGY553 (ATCC). Transformed yeast was selected on Edinburgh minimal medium (EMM)–Ura plates for pNCH1472-myc-sty1, EMM–Leu for pSGP73-HA-snr1, and EMM–Ura–Leu for double transformation.

2.4.16 Co-immunoprecipitation and western blotting

Transformed yeast (KGY553) containing pNCH1472-myc-sty1 or pSGP73-HA-snr1 or both were cultured overnight in 10mL EMM selection medium. Yeast pellets were washed in 5mL of cold TE buffer before protein extraction. To lyse cells, 1mL of lysis buffer (50mM Tris-

HCl pH 7.5, 0.2% Tergitol, 150mM NaCl, 5mM EDTA, Complete Protease Inhibitor tablet) and 600μL glass beads were added to each tube and mixed in a beater for two rounds of 10 minutes each. Protein extracts were centrifuged for 10 minutes at 14,000 rpm at 4°C. Then, 500μL of supernatant was immunoprecipitated overnight using 20μL of EZview™ Red Anti-c-Myc Affinity Gel (Sigma-Aldrich E6654) or EZview™ Red Anti-HA Affinity Gel (Sigma-Aldrich E6779). Primary antibodies used in our analysis were anti-c-Myc (Santa Cruz sc-789), anti-HA (Roche 12CA5), and anti-γ-tubulin (Sigma-Aldrich T5192).

2.4.17 Construction of yeast deletion strains

The *snr1*Δ strain was obtained from the Bioneer *Schizosaccharomyces pombe* Genome-wide Deletion Library. The deletion strain was verified by PCR using primers SpEhd3_Up_Fwd and Sp_Dn_Rev spanning the 3' end of *snr1* and the region immediately downstream. Primers specific for KanMX4 (KanMX4-Fwd and KanMX4-Rev) were used to detect the deletion cassette. A PCR-based strategy was used to construct the *sty1*Δ strain. Briefly, in the first round of PCR, primers (PFA6a_Sty1_Fwd and PFA6a_Sty1_Rev) with 20 base pairs (bp) homology to the regions upstream and downstream of sty1, respectively, were synthesized for PCR of the pFA6a-KanMX6 cassette. Primers with 20 bp homology to the pFA6a-KanMX6 were synthesized to PCR 290 bp upstream (Sty1Del-Up_Fwd and Sty1Del-Up_Rev) and 290 bp downstream (Sty1Del-Dn_Fwd and Sty1Del-Dn_Rev) of sty1, not including the sty1 gene. The three PCR products were stitched together sequentially with a second round of PCR. Stitch PCR of the upstream region and pFA6a-KanMX6 and of the downstream region and pFA6a-KanMX6 were carried out separately. In the third round of PCR, both upstream and downstream stitched PCR products were further stitched together to produce a final product of pFA6a-KanMX6 flanked on the 5' and 3' ends by 290 bp that are homologous to the upstream and downstream

regions of chromosomal sty1 (Sty1Del-Up_Fwd and Sty1Del-Dn_Rev). The final PCR product was transformed into *S. pombe* 972h- canonical wild-type (ATCC). Transformed yeast was selected on yeast extract with supplements (YES) media plates containing 150mg/L G418. Insertion of the pFA6a-KanMX6 cassette by homologous recombination at the sty1 locus was verified by PCR using primers to target the entire cassette (Sty1Del-Up_Fwd and Sty1Del-Dn_Rev) and to target a sty1 internal region of 401 bp (Sty1_Fwd and Sty1_Rev). Sequences of primers used for deletion and verification of strains in this study are listed in Table 2.1.

2.4.18 Stress Sensitivity Assays

S. cerevisiae was grown in YEPD medium and *S. pombe* was grown in YES medium. All yeast strains were initially grown as a starter culture overnight at 30°C. From the starter culture, yeast cells were diluted into fresh medium to an initial OD_{600nm} = 0.2. The cultures were grown to mid-log phase (OD_{600nm} = 0.7). The *S. cerevisiae* and *S. pombe* strains were serially diluted 4-fold in sterile water and spotted onto YEPD and YES plates, respectively, containing various stressors. Spotted plates were incubated at 30°C and yeast growth was assessed after 3 days.

2.5 RESULTS

2.5.1 Comparison of known interactions in *S. cerevisiae* and *S. pombe*

The number of known protein-protein interactions in *S. pombe* is disproportionately lower than in other model eukaryotic organisms and human. The number of all known interactions in *S. cerevisiae* and *S. pombe* were estimated by assimilating seven commonly-used

Primer Name	Primer Sequences (5'-3')
Sty1-pNCH1472-Forward	AAGGAAAAAAGCGGCCGCATGGCAGAATTTATTCGTAC
Sty1-pNCH1472-Reverse	GGTGTCGACGGATTGCAGTTCATTATCCATG
snr1-pSGP73-Forward	AAGGAAAAAAGCGGCCGCATGGGATTGAAATTAAATATC
snr1-pSGP73-Reverse	GGAAGATCTCTATAAATAAGGATAAGTC
SpEhd3_Up_Fwd	CTTAAACAGCCTGATTTTGT
SpEhd3_Dn_Rev	AACTATCGTACGCACAGCTA
KanMX4-Fwd	TTAGCTTGCCTCGTCCCC
KanMX4-Rev	TTTCGACACTGGATGGCG
Sty1Del-Up_Fwd	TACAAGCAAACACCACAATC
Sty1Del-Up_Rev	TTAATTAACCCGGGGATCCGTTTATTCAAACTGGTTACAAAAAGGAC
PFA6a_Sty1_Fwd	TTGTAACCAGTTTGAATAAACGGATCCCCGGGTTAATTAA
PFA6a_Sty1_Rev	AGGCTTTATCTACAACTTGTGAATTCGAGCTCGTTTAAAC
Sty1Del-Dn_Fwd	GTTTAAACGAGCTCGAATTCACAAGTTGTAGATAAAGCCTTAAAAGTTGTTC
Sty1Del-Dn_Rev	ACACCACACTTGAAAATCGC
Sty1_Fwd	AATTGAGACGATTTGCAGTAAAAAC
Sty1_Rev	TAATACGCTTACGAGGATCAAAGAC

Table 2.1 Primers Used

databases – BioGRID (Stark et al., 2011), DIP (Salwinski et al., 2004), IntAct (Kerrien et al., 2012), iRefWeb (Turner et al., 2010), MINT (Ceol et al., 2010), MIPS (Mewes et al., 2011), and VisANT (Hu et al., 2007). There are 110,443 known interactions for budding yeast, but only 4,038 for fission yeast. Furthermore, previous studies have shown that only those interactions or interaction sets that have been validated by at least two independent assays are of high quality (Cusick et al., 2009; Das et al., 2012). Based on this criterion, 519 fission yeast interactions are of high quality, as opposed to 25,335 high-quality interactions known in budding yeast. Of these, only 160 *S. pombe* interactions are binary (a direct biophysical interaction between the two proteins), as opposed to 11,936 in *S. cerevisiae.* These numbers (Table 2.2) indicate the extent to which the fission yeast interactions are underexplored and necessitate the systematic mapping of its interactome network.

2.5.2 StressNet: a large-scale high-quality protein interactome network for stress response and cellular signaling in *S. pombe*

The subset of 658 genes for this study was selected using Gene Ontology (GO) (Ashburner et al., 2000) biological process (BP) functional annotations for fission yeast (Figure 2.1A and Materials and Methods). To generate a high-quality high-coverage stress-response interactome map for *S. pombe*, we screened all possible protein pairs (>430,000) in this space three times using an improved version of the high-quality HT-Y2H system, as we had done for *S. cerevisiae* (Yu et al., 2008) (Materials and Methods). We further developed our "Stitch-seq" method to leverage the power of paired-end Illumina sequencing, which greatly increases throughput and sequencing depth with much decreased costs (Yu et al., 2011). The resulting protein interactome network, StressNet (Figure 1.1B), comprises 235 high-quality binary interactions among 200 proteins. Of these, 218 interactions were previously unknown. To

validate our experimental pipeline and the quality of StressNet, we selected a set of 54 well-documented protein interactions from the literature ["positive reference set" (PRS); Materials and Methods] and 43 random protein pairs that have never been reported or predicted to interact ["random reference set" (RRS); Materials and Methods]. To construct the PRS, we selected interactions from the 160 high-quality binary interactions, where both interacting proteins are in our search space. These 54 PRS interactions are supported by at least two independent assays from two different publications. The 43 RRS pairs were generated from a random selection of all possible protein pairs within our search space for which no interaction has yet been detected by any method. Furthermore, since fission yeast interactions are considerably underexplored, we also required that their corresponding budding yeast ortholog pairs have never been reported to interact. 20 PRS interactions were successfully confirmed in our pipeline, whereas none of the RRS pairs were detected as positives (Figure 2.1C). Therefore, the sensitivity [fraction of true positives among all actual positives (Yu et al., 2008)] of our Y2H assay is 37.0%.

To directly measure the quality of our Y2H-identified interactions (Braun et al., 2008; Yu et al., 2008), all 235 interactions detected in our HT-Y2H screen were systematically re-tested by two orthogonal assays: the protein complementation assay (PCA) (Materials and Methods) (Remy and Michnick, 2006) and the well-based nucleic acid programmable protein array (wNAPPA) (Materials and Methods) (Ramachandran et al., 2004), producing a fully-verified large-scale interactome map. The confirmation rates of our interactions with both orthogonal assays are very similar to those of the PRS, further validating the high quality of StressNet (Braun et al., 2008; Yu et al., 2008) (Figure 2.1C). Moreover, the precision [fraction of true positives among all assay-reported positives (Yu et al., 2008)] of our interactome is measured at 95.3% (Materials and Methods). To assign a confidence score to each interaction in our network,

we implemented a random forest algorithm to integrate results from the three orthogonal assays (Figure 2.2A and Materials and Methods). Finally, to evaluate the topological properties of our network, we plotted the degree (number of interactions each protein has) distribution of StressNet (Figure 2.1D). It is known that protein interactomes are small-world scale-free networks (Barabasi and Albert, 1999; Jeong et al., 2000). Our new interactome has topological properties similar to other large-scale biological networks.

To assess the biological relevance of this network, we investigated overall relationships between protein pairs using expression and genetic interaction profile similarities (Roguev et al., 2008; Rustici et al., 2004), cellular co-localization (Matsuyama et al., 2006), and GO functional similarities (Ashburner et al., 2000). We found significant enrichment of interactions in StressNet whose proteins are colocalized and whose corresponding gene pairs are co-expressed, of similar genetic interaction profiles, and are functionally similar (all $P < 0.05$ using a cumulative binomial test; Figures 2.3A-D), relative to random expectation. Furthermore, the enrichment of StressNet in all four categories is the same as that of high-quality literature-curated binary interactions. These results confirmed the high quality of StressNet. More importantly, our results show that these interactions are likely to be functionally relevant.

2.5.3 Evolutionary relationships in StressNet

For biological networks, evolutionary relationships are commonly measured in terms of conservation and rewiring: If a pair of interacting proteins in one species has corresponding orthologs in another that also interact, then the interaction is considered to be conserved (an

Interaction type	Number of interactions (*S.c.*)	Number of interactions (*S.p.*)
All	110,443	4,038
HQ	25,335	519
Binary All	18,973	1,059
Binary HQ	11,936	160

Table 2.2 Interactome Sizes

S.c., *Saccharomyces cerevisiae*; *S.p.*, *Schizosaccharomyces pombe*

interolog); otherwise, the interaction is considered to be rewired (Matthews et al., 2001; Shou et al., 2011; Yu et al., 2004) (Figure 2.4A). To understand key principles governing the evolution of protein-protein interactions, especially for those in stress response and signaling pathways, we compared the interactions in StressNet to their corresponding ortholog pairs in *S. cerevisiae*. We experimentally tested all corresponding *S. cerevisiae* protein pairs of the 235 interactions in StressNet and found that for 35 interactions, the corresponding budding yeast ortholog pairs were detected as interacting by our Y2H experiments. We developed a Bayesian framework to calculate the percentage of conserved interactions based on three parameters—the proportion of observed conserved interactions (35/235 = 14.9%) and the precision (95.3 ± 4.7%) and the sensitivity (37.0 ± 4.4%) of our Y2H assay. Substituting appropriate values, the percentage of conserved interactions between *S. pombe* and *S. cerevisiae* is calculated as 36.3 ± 2.9% (Figure 1.4B).

Using an orthogonal approach, we supplemented *S. cerevisiae* interactions detected in our Y2H experiments with high-quality known *S. cerevisiae* interactions curated from the literature to obtain 55 more StressNet interactions for which the corresponding budding yeast orthologs were reported to interact in the literature (Das and Yu, 2012). There are 90 (35 + 55) conserved interactions in total, and the conservation is 38.3 ± 3.2%, consistent with the conservation calculated using the Bayesian framework (Figure 2.4B). Furthermore, this agreement shows that after combining our Y2H experimental results with high-quality literature-curated interactions, the number of known interactions in our search space in *S. cerevisiae* is nearly complete, because if there were still a large number of unidentified interactions, the observed proportion of conserved interactions based on literature-curated interactions would have been much lower.

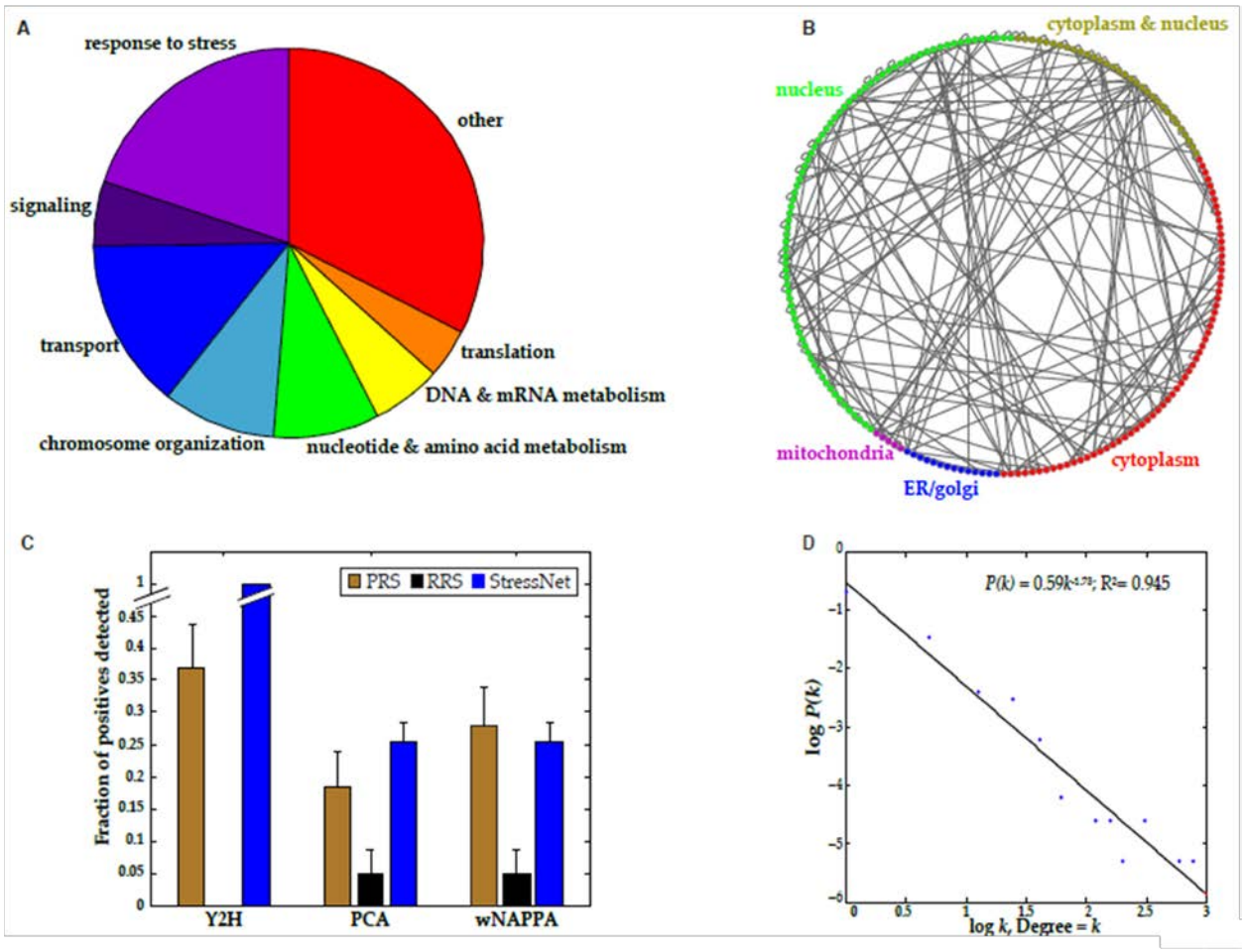
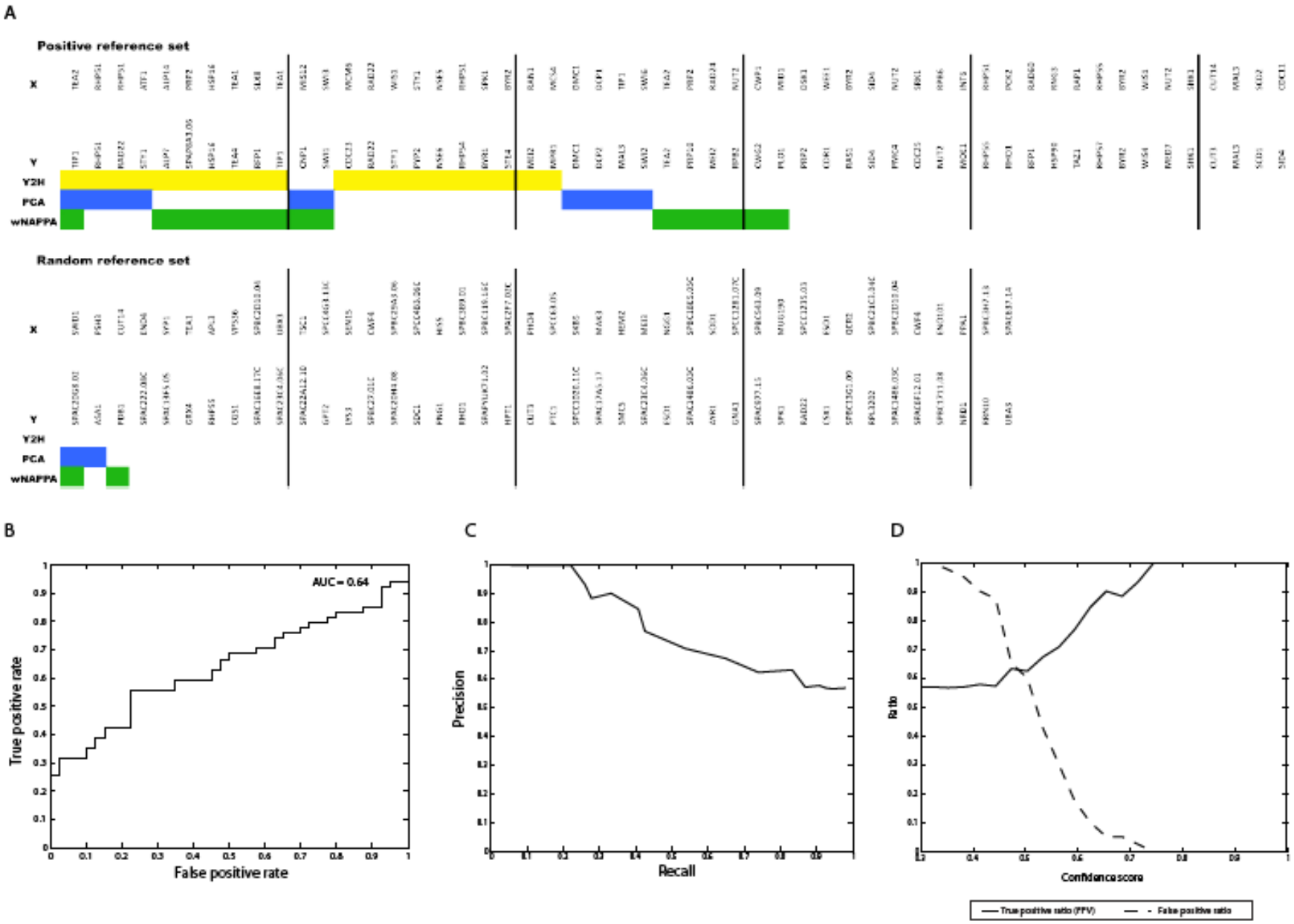


Figure 2.1 *S. pombe* Stress Response Binary Interactome Network, StressNet

(A) Functional classification of the proteins included in our high-quality, high-coverage HT-Y2H screen. (B) Network view of the stress response binary interactome network in *S. pombe*. (C) Fractions of protein pairs in PRS, RRS, and StressNet that tested positive using Y2H, PCA, and wNAPPA. Data are shown as measurements + statistical error (SE). (D) Degree distribution of StressNet. $P(k)$ is the probability that a protein has a degree = k.

Figure 2.2 Orthogonal Validations of StressNet

(A) The colors yellow, blue, and green correspond to detection by Y2H, PCA, and wNAPPA respectively. The vertical lines are visual dividers to enhance readability. (B) Receiver operating characteristic (ROC) curve illustrating the performance of the random forest classifier (evaluated by five-fold cross-validation). (C) Precision-recall curve illustrating the performance of the random forest classifier (evaluated by five-fold cross-validation). (D) Distribution of true positive ratio (PPV) and false positive ratio with confidence score.

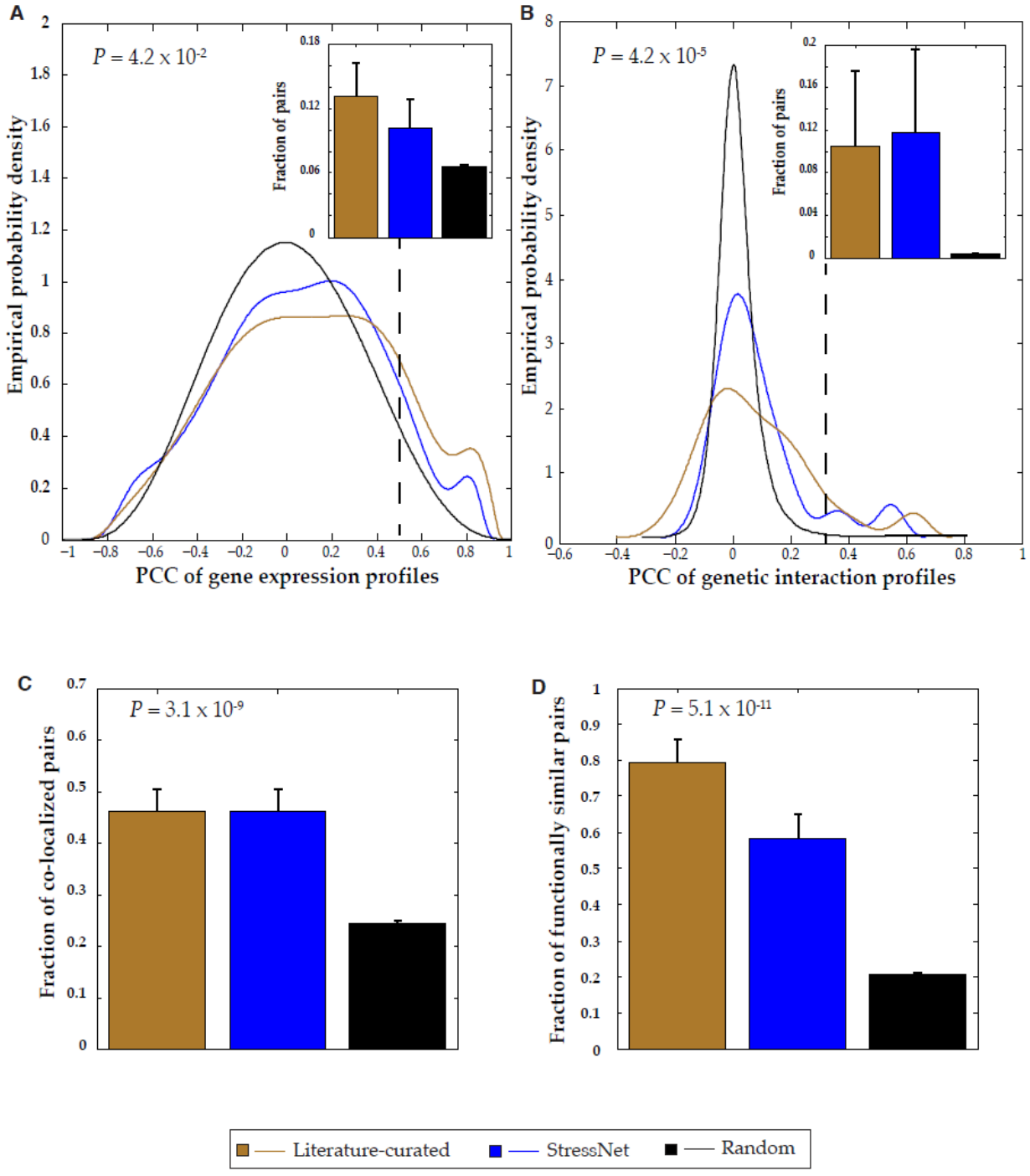

A
Positive reference set
X
Y
Y2H
PCA
wNAPPA
Random reference set
X
Y
Y2H
PCA
wNAPPA
B
AUC = 0.64
True positive rate
False positive rate
C
Precision
Recall
D
Ratio
Confidence score
True positive ratio (PPV)
False positive ratio

Because it is always difficult to determine a negative interaction (Ben-Hur and Noble, 2006; Yu et al., 2008), to ensure the set of rewired interactions is of high quality, we used a stringent set of criteria to define them as those StressNet interactions without corresponding *S. cerevisiae* ortholog pairs and those interactions whose corresponding *S. cerevisiae* ortholog pairs have other high-quality interactions but have never been reported as interacting in the literature or tested positive in our Y2H experiments, and these ortholog pairs are known to have different cellular localizations (Huh et al., 2003).

Proteins encoded by essential genes, those when deleted cause lethality, tend to have more interacting partners (hubs) and also evolve more slowly than nonessential ones (Fraser et al., 2002; Hirsh and Fraser, 2001). We found that essential and nonessential genes (Kim et al., 2010) in our interactome were equally likely to be involved in conserved interactions (Figure 2.4C), contrary to previous studies (Fraser et al., 2002). Stress response and signal transduction pathways play a crucial role in the process of adaptation to distinct ecological environments. As measured by the ratio of nonsynonymous to synonymous substitution rates (*dN/dS*) (Nei and Gojobori, 1986; Rhind et al., 2011), we found that the essential genes in these pathways evolve at the same rate as the nonessential genes in the pathways evolve, although, on average, all essential genes in the genome evolve significantly slower (Figures 2.5A and B) than do nonessential genes. To ensure that this is not an artifact of the calculation method, we also calculated the *dN/dS* values for all essential and nonessential genes. Consistent with earlier findings (Das and Yu, 2012), we observed that, overall, the essential genes had a significantly lower average *dN/dS* (Figures 2.5A and B). The average *dN/dS* for all stress response genes is not significantly different from that for the entire genome (Figure 2.5C). The *dN/dS* distributions for these two species are highly similar (Figure 2.5D). This finding is consistent with analyses

Figure 2.3 Biological Properties of StressNet Interactions

(A) *PCC* distribution of expression profiles of interacting and random protein pairs (dashed line corresponds to *PCC* cutoff above which pairs are considered to be significantly coexpressed; inset shows the fraction of significantly coexpressed pairs). (B) *PCC* distribution of genetic interaction profiles of interacting and random protein pairs (dashed line corresponds to *PCC* cutoff above which pairs are considered to be significantly similar; inset shows the fraction of pairs with significantly similar interaction profiles). (C) Enrichment of colocalized protein pairs. (D) Enrichment of protein pairs sharing similar functions. For each panel, the random set is constructed by considering all pairwise combinations of genes or proteins in the corresponding space. All *P* values represent comparisons between StressNet interactions and random pairs using a cumulative binomial test. Inset graphs and data in (C) and (D) are shown as measurements + SE.


A
P = 4.2 x 10-2
Empirical probability density
PCC of gene expression profiles
Fraction of pairs
B
P = 4.2 x 10-5
Empirical probability density
PCC of genetic interaction profiles
Fraction of pairs
C
P = 3.1 x 10-9
Fraction of co-localized pairs
D
P = 5.1 x 10-11
Fraction of functionally similar pairs
Literature-curated
StressNet
Random

that suggest that these species are at comparable evolutionary distances from *S. pombe* (Hu et al., 2007; Mewes et al., 2011) and confirm that there are no inherent biases in our *dN/dS* calculations. Thus, our findings suggest that essential genes in stress response and signal transduction pathways are under less negative selection such that their interactions are rewired for adaptive advantages through evolution.

To better understand the mechanisms underlying conservation and rewiring of interactions, we examined the relationship between sequence similarity of orthologous pairs and interaction conservation rates. Consistent with expectation (Yu et al., 2004), interactions involving proteins with higher overall sequence similarity or identity were more likely to be conserved (Figures 2.4D and 2.5E). However, proteins interact through specific domains (Finn et al., 2010); therefore, we examined the role of sequence similarity of these interfaces in determining the conservation of corresponding interactions. Previous studies have established a homology modeling approach (Kim et al., 2006; Wang et al., 2012) to locate interaction interfaces using cocrystal structures in the Protein Data Bank (Berman et al., 2000) and have found that analysis of these interfaces provides insights into their evolutionary rate (Kim et al., 2006). The conservation of an interaction depends on the conservation of the interfaces involved (Espadaler et al., 2005). Using a similar approach, we inferred interaction interfaces for proteins involved in 161 interactions in our network (Materials and Methods). We found no significant correlation between the similarity or identity of interaction interfaces and the conservation of the corresponding interactions (Figures 2.4E and 2.5F). Examination of the average *dN/dS* ratios for proteins with different numbers of rewired interactions showed that the selection pressure on the gene did not affect the degree to which the interactions of the corresponding protein were rewired (Figure 2.4F), further indicating that the rewiring of interactome networks and the

divergence of related species are not completely dictated by evolution detected at the sequence level.

2.5.4 Functional profile of conserved and rewired interactions

To investigate whether gene pairs encoding proteins involved in conserved and rewired interactions are differently regulated at the transcriptional level, we measured global coexpression between these pairs using the *PCC*. Global coexpression means that the patterns of gene expression of both genes are the same (Figure 2.6). Whereas conserved interactions had the highest fraction of coexpressed pairs, gene pairs encoding proteins involved in rewired interactions were also significantly more coexpressed than random in *S. pombe* (Figure 2.7A). We also calculated coexpression relationships for the corresponding budding yeast pairs. By definition, the conserved pairs also interact in budding yeast, but the rewired pairs do not. The enrichment in gene expression is consistent with this distinction: Gene pairs encoding proteins involved in conserved interactions were coexpressed, and genes encoding rewired pairs were not significantly more enriched than random expectation in *S. cerevisiae* (Figure 2.7A).

PCC captures only global coexpression relationships but cannot capture local or transient coexpression that occurs only under certain conditions (Figure 2.6). Stable and transient interactions both have important biological functions—the former constitute tightly connected modules, whereas the latter form key links between modules, especially in signal transduction pathways, and are more important than the stable ones or random interactions in maintaining the integrity of cellular networks (Das et al., 2012). To detect transient interactions, we used the local expression-correlation scores (LES) (Das et al., 2012; Qian et al., 2001). Rewired interactions in fission yeast had significantly higher LES values (Figure 2.7B) than both

Figure 2.4 Evolutionary Analysis of Interactions

(A) Schematic of conserved and rewired interactions between the two yeast species. *S.p.*, *S. pombe*; *S.c.*, *S. cerevisiae*. (B) Conservation rate (fraction of conserved interactions) in our interactome calculated in two different ways. Measured represents the value calculated using a Bayesian framework that incorporates the precision and recall of our assay. Literature represents the value estimated using budding yeast interactions reported in the literature. (C) Fractions of conserved interactions involving essential and nonessential proteins. The differences in (B) and (C) are not significant based on a cumulative binomial test. (D) Distribution of the fraction of conserved interactions as a function of overall sequence similarity. (E) Distribution of the fraction of conserved interactions as a function of sequence similarity of interaction interfaces.

For (D) and (E), *P* values are used to test whether there is a significant difference (using a cumulative binomial test) in conservation percentage between the groups corresponding to the lowest and highest similarity percentages. R^2 (coefficient of determination) represents the significance of the correlation between conservation and similarity percentages. (F) Distribution of *dN/dS* [ratio of the number of nonsynonymous substitutions per nonsynonymous site (Costanzo et al.) to the number of synonymous substitutions per synonymous site (*dS*)] as a function of the number of rewired interactions. Differences are not significant as determined by a two-sided Kolmogorov-Smirnov test. Data are shown as measurements + SE.

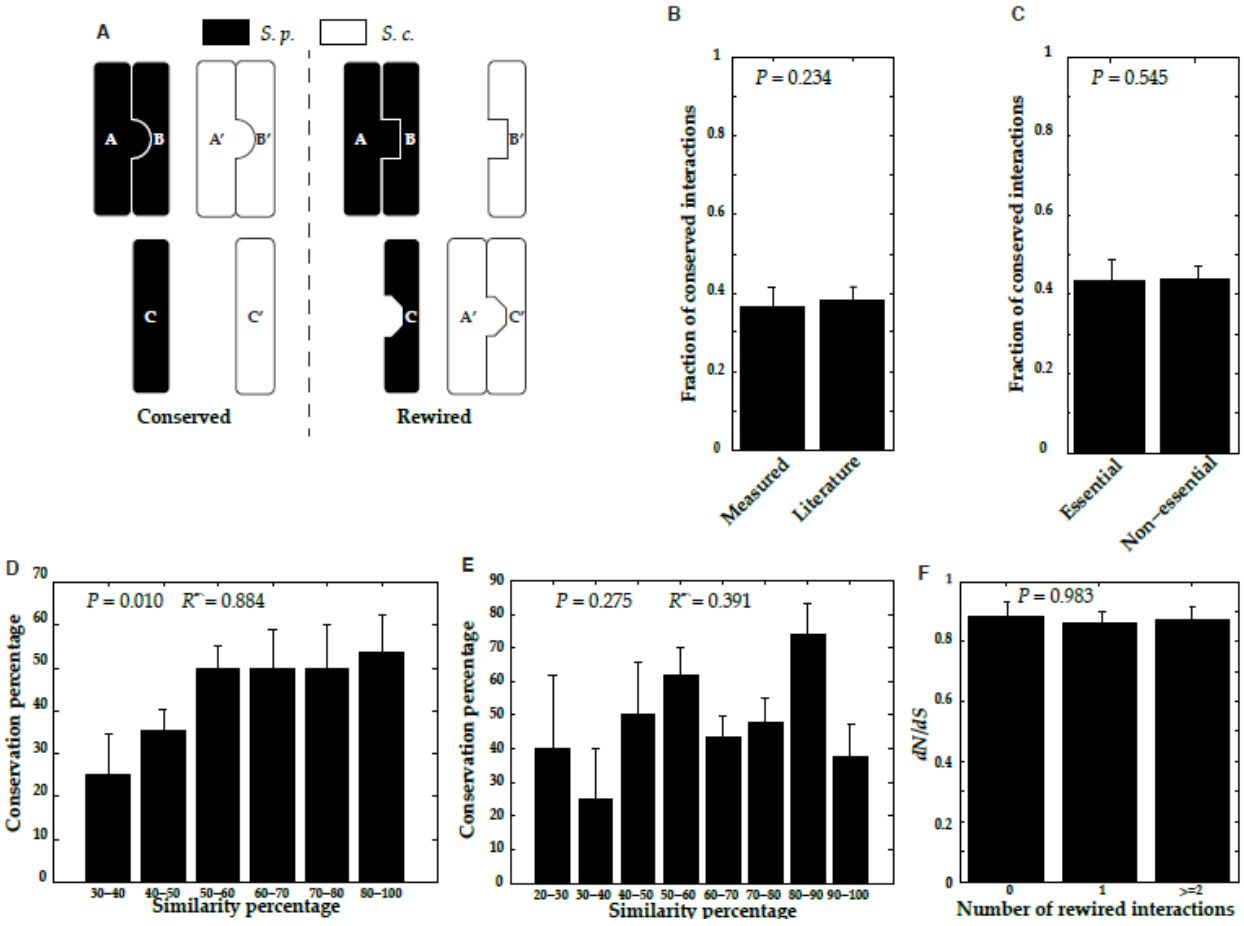
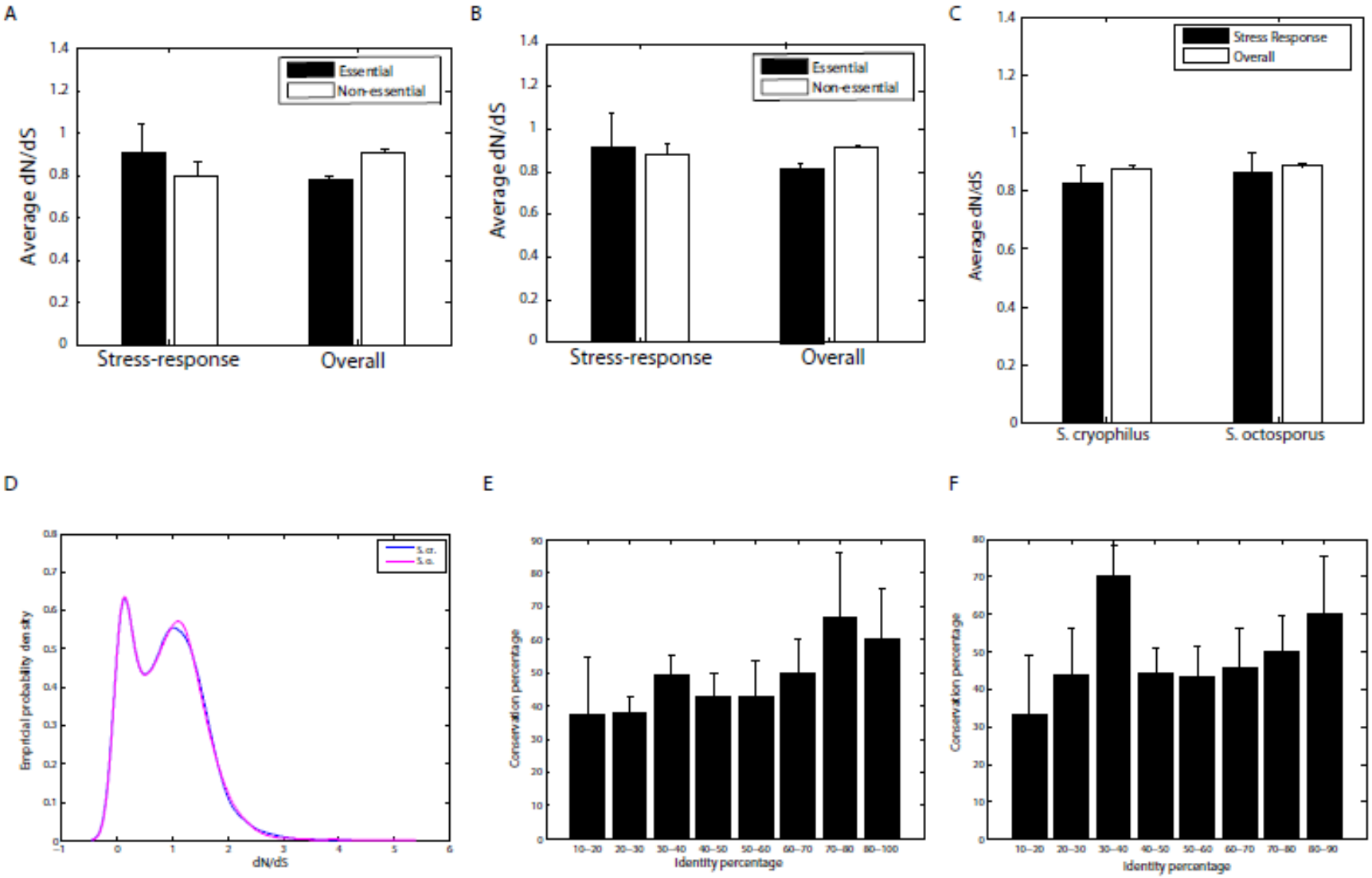

A
S. p.
S. c.
A
B
A′
B′
C
C′
Conserved
Rewired
B
P = 0.234
Fraction of conserved interactions
Measured
Literature
C
P = 0.545
Fraction of conserved interactions
Essential
Non−essential
D
P = 0.010
R² = 0.884
Conservation percentage
30–40
40–50
50–60
60–70
70–80
80–100
Similarity percentage
E
P = 0.275
R² = 0.391
Conservation percentage
20–30
30–40
40–50
50–60
60–70
70–80
80–90
90–100
Similarity percentage
F
P = 0.983
dN/dS
0
1
>=2
Number of rewired interactions

Figure 2.5 Comparison of Evolutionary Rates of Genes

(A) *dN/dS* values for essential and nonessential *S. pombe* genes in stress-response pathways and overall calculated using orthologs from *S. cryophilus.* (B) *dN/dS* values for essential and non-essential *S. pombe* genes in stress-response pathways and overall calculated using orthologs from *S. octosporus.* (C) Average *dN/dS* values for *S. pombe* genes in stress-response pathways and overall calculated using orthologs from *S. cryophilus* and *S. octosporus.* (D) Distribution of *dN/dS* values for *S. pombe* genes calculated using orthologs from *S. cryophilus* and *S. octosporus*. *S. cryophilius* is denoted by *S.cr.* and *S. octosporous* denoted by *S.o.* (E) Distribution of conservation percentage across overall identity percentage. (F) Distribution of conservation percentage across identity of interacting domains.


A
Average dN/dS
Essential
Non-essential
Stress-response
Overall
B
Average dN/dS
Essential
Non-essential
Stress-response
Overall
C
Average dN/dS
Stress Response
Overall
S. cryophilus
S. octosporus
D
Empirical probability density
dN/dS
E
Conservation percentage
Identity percentage
10–20
20–30
30–40
40–50
50–60
60–70
70–80
80–100
F
Conservation percentage
Identity percentage
10–20
20–30
30–40
40–50
50–60
60–70
70–80
80–90

conserved interactions and random expectation, suggesting that transient interactions are more likely to be rewired through evolution. Rewired pairs in budding yeast had lower LES values than random pairs (Figure 2.7B), indicating that gene regulation for these pairs is also rewired.

Next, we examined the GO functional similarities between interacting proteins involved in conserved and rewired interactions. Whereas conserved interactions had higher functional similarity than rewired interactions in fission and budding yeast, interacting protein pairs in both categories were significantly more functionally similar than random (Figure 2.7C). This is in agreement with previous findings that conserved interactions tend to be in modules with specific functions, whereas rewired interactions tend to be intermodular and have greater diversity in functions, whereas rewired interactions tend to be intermodular and have greater diversity in function (Das et al., 2012).

2.5.5 Novel modes of rewiring uncovered by cross-species interactome mapping

To further understand the meaning of "conservation" of interactions and experimentally explore the molecular mechanisms through which interaction interfaces evolve, we performed a systematic cross-species interactome mapping. Using orthologous pairs of interacting proteins in *S. cerevisiae* and *S. pombe*, we examined whether a protein in one species interacted with the ortholog of its partner in the other (Figure 2.8A). Because we could detect the original interacting pairs from the same species with our Y2H experiments, we know that all four proteins are correctly expressed, folded, and amenable to detection by our Y2H approach, thereby avoiding technical false negatives. The traditional definition of conservation implies the notion of conserved interfaces across different species. Here, the interface of a conserved

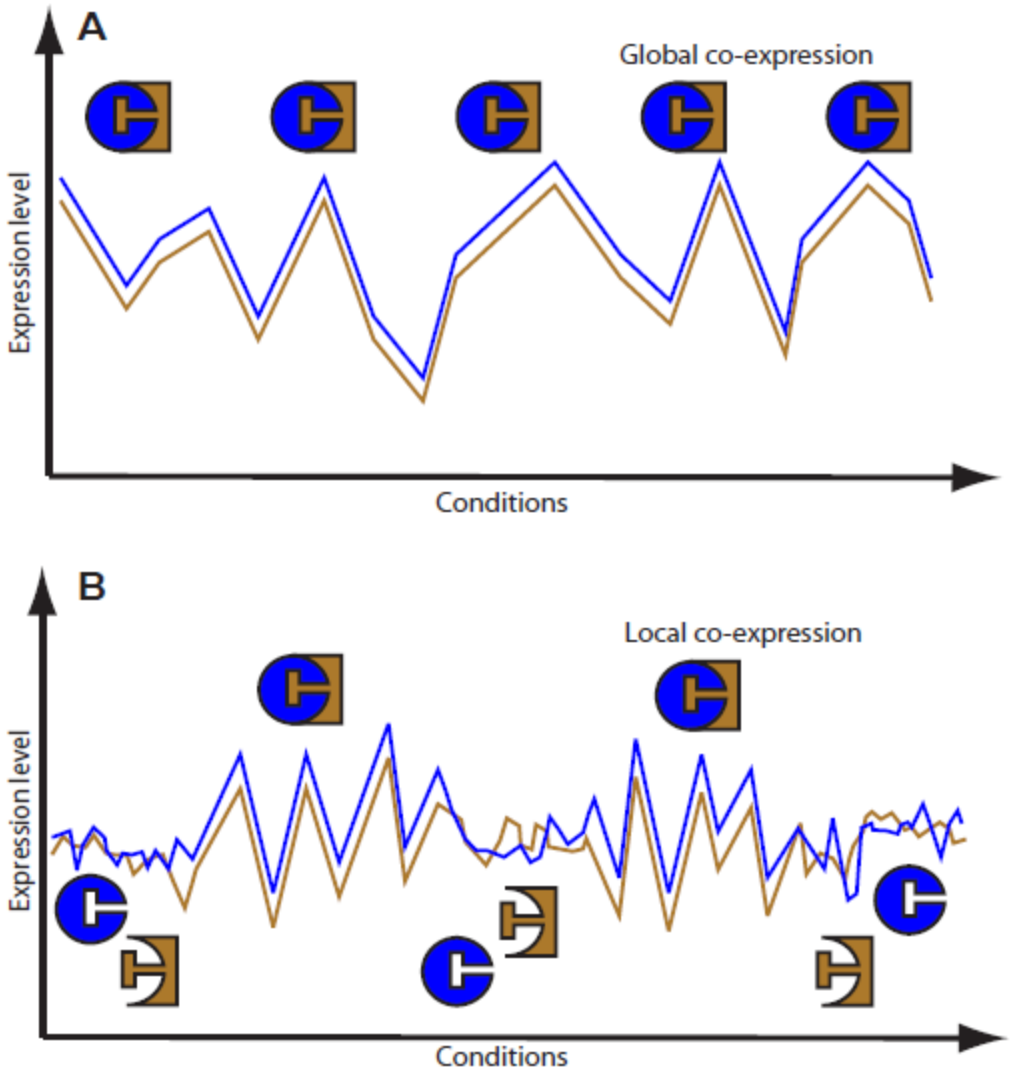


Figure 2.6 Graphical Representation of Global and Local Coexpression of Genes that Express Interacting Proteins

(**A**) Global coexpression showing synchronous expression of the two genes and consistent interaction of the products. (**B**) Local coexpression showing desynchronized expression and interaction of the products only during periods of synchronous expression. Shapes represent two interacting proteins. Condition refers to different time points corresponding to which expression values are measured. [Adapted from (Das et al., 2012)].

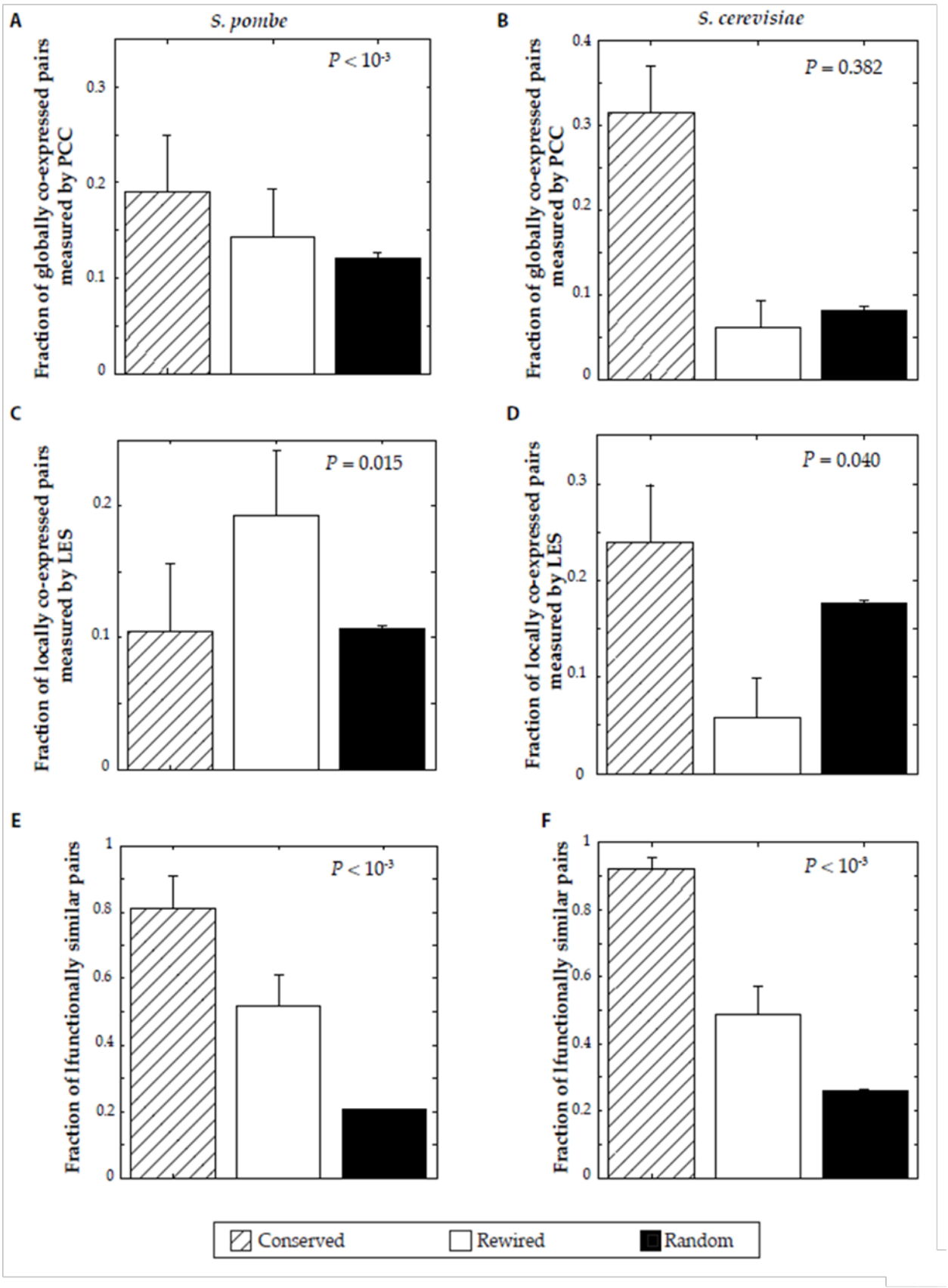
interaction in fission yeast is considered “intact” if the proteins involved could also interact with the corresponding orthologs of their partners in budding yeast; otherwise, the interface is considered “coevolved” (Figure 2.8A). We found that these conserved interactions equally likely result from an intact interface or coevolved interface that formed new interaction interfaces that were unrecognizable by their orthologous counterparts in the other species (Figure 2.8B). Earlier studies have suggested that interacting proteins may coevolve to maintain structural complementarity and binding specificity (Goh et al., 2000; Hakes et al., 2007a; Kim et al., 2004). In this calculation, we used a lenient definition for an intact interface: We considered the interface intact if one or both of the cross-species interactions were positive, which provides a lower bound estimation of coevolution between interacting proteins.

2.5.6 Divergence of the Sty1 stress-response pathway through interaction conservation and rewiring

In *S. pombe*, Sty1 is activated in response to various stresses, including oxidative and osmotic stress, starvation, and other conditions (Gasch, 2007; Shiozaki and Russell, 1996). Sty1 has orthologs in *S. cerevisiae* (Hog1, with 89% sequence similarity) and human (p38, with 69% sequence similarity). Both p38 and Sty1 respond to a wide range of stresses, and both are different from Hog1 in terms of function (Bone et al., 1998). With our stress response interactome, we detected key interactions at every step of the MAPK signal transduction pathway and, therefore, completely recapitulated the entire Sty1 pathway (Figure 2.9A). This confirmed the sensitivity and accuracy of our HT-Y2H method, especially for discovering transient interactions in signaling pathways. Among all Sty1 interactions in StressNet, those with its activator (Wis1) and inhibitor (Pyp2) were both conserved between the two yeast species, and

Figure 2.7 Functional Analysis of Conserved and Rewired Interactions in *S. pombe* and *S. cerevisiae*

(A) Fractions of globally coexpressed pairs (as measured by *PCC*) among conserved and rewired interactions. (B) Fractions of locally coexpressed pairs (as measured by LES) among conserved and rewired interactions. (C) Fractions of functionally similar pairs among conserved and rewired interactions. For each panel, the random set is constructed by considering all pairwise combinations of genes/proteins in the corresponding space. All *P* values represent comparisons between rewired interactions and random pairs using a cumulative binomial test. Data are shown as measurements + SE.


A
S. pombe
P < 10-3
Fraction of globally co-expressed pairs measured by PCC
0
0.1
0.2
0.3
B
S. cerevisiae
P = 0.382
Fraction of globally co-expressed pairs measured by PCC
0
0.1
0.2
0.3
0.4
C
P = 0.015
Fraction of locally co-expressed pairs measured by LES
0
0.1
0.2
D
P = 0.040
Fraction of locally co-expressed pairs measured by LES
0
0.1
0.2
0.3
E
P < 10-3
Fraction of functionally similar pairs
0
0.2
0.4
0.6
0.8
1
F
P < 10-3
Fraction of functionally similar pairs
0
0.2
0.4
0.6
0.8
1
Conserved
Rewired
Random

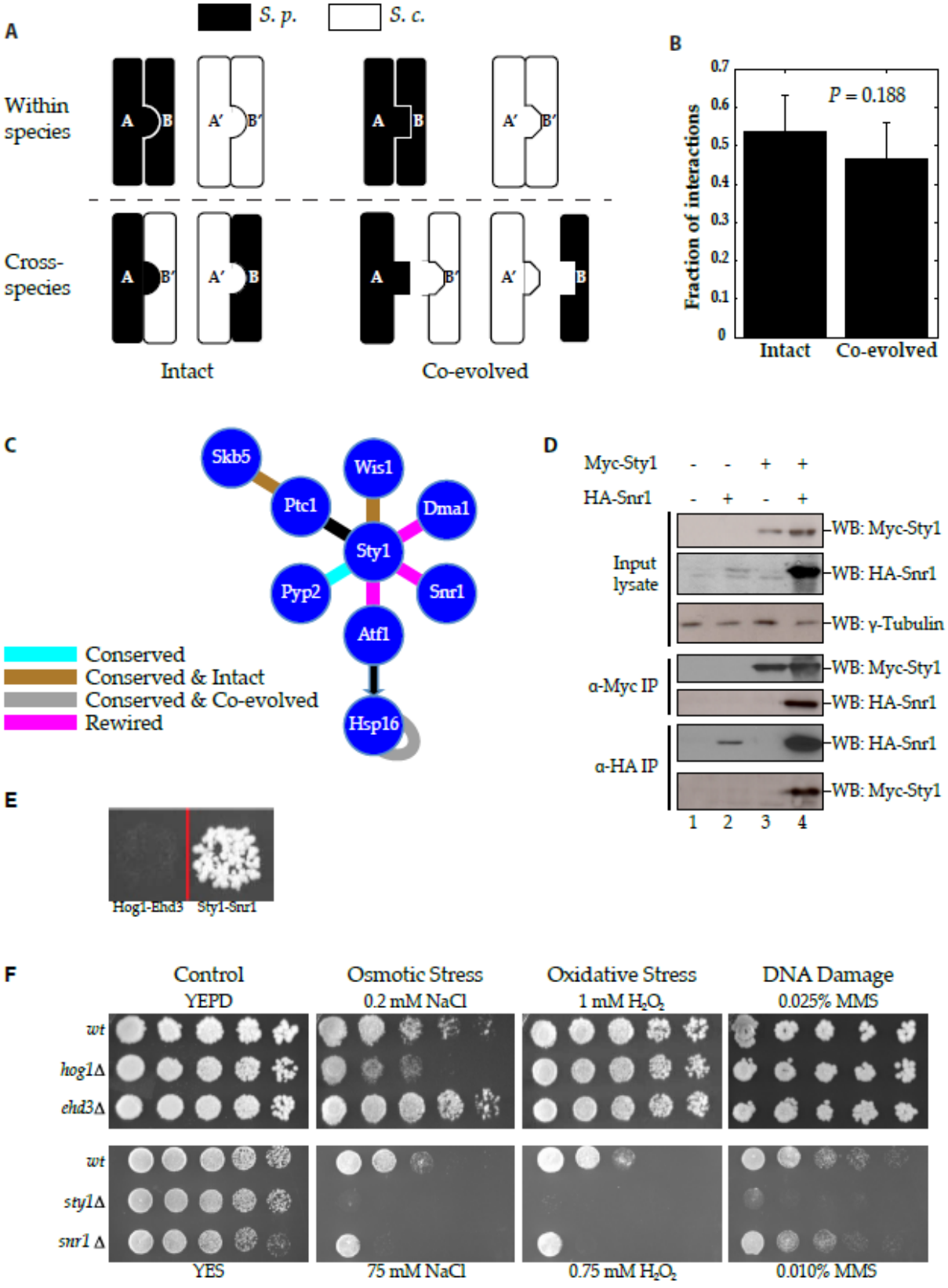
the Sty1-Wis1 interaction interface was intact. By contrast, the interaction between Sty1 and its known target in fission yeast, Atf1, represented a rewired interaction (Figure 2.8C). We also identified a previously unknown interactor of Sty1: SPBC2D10.09, a protein that we named Snr1 (Sty1- interacting stress response protein). To confirm this interaction *in vivo*, we performed coimmunoprecipitation of tagged proteins expressed in *S. pombe* (Figure 2.8D). The amount of Snr1 pulled down in the presence of Sty1 was greater than that pulled down in the absence of Sty1, indicating that the interaction with Sty1 stabilizes Snr1 (Figure 2.8D). The corresponding orthologous pair of Hog1 and Ehd3 in *S. cerevisiae* did not interact by Y2H (Figure 2.8E). Cells lacking *snr1* (*snr1Δ* cells) grew slower under stress, similar to *sty1Δ* cells (Figures 2.8F and 2.9B), whereas the growth of *ehd3Δ* cells was not compromised. These results suggested that Snr1 is a component of the Sty1 pathway and that its functions diverged from its budding yeast counterpart. Moreover, snr1 also has a human ortholog, HIBCH, further investigation of which may expand our knowledge of the human p38 MAPK pathway.

2.6 CONCLUSIONS AND DISCUSSION

We generated StressNet—a high-quality, high-coverage binary interactome for stress response and signal transduction pathways in the fission yeast *S. pombe*. All interactions were verified by three orthogonal assays and assigned probabilistic confidence scores. We performed comparative network analysis to study the evolution of protein interactomes between the fission and budding yeast species. Although 84% of StressNet interactions have corresponding orthologous pairs in *S. cerevisiae*, only about 40% of these interactions are conserved, indicating considerable evolutionary changes beyond simple sequence orthology. Thus, the interolog concept should be used with caution to infer interactions across species, especially if the two are

Figure 2.8 Analysis of Intact and Coevolved Interactions

(A) Schematic of intact and coevolved interactions. (B) Fractions of intact and coevolved interactions in our interactome. Data are shown as measurements + SE. No significant difference detected by a cumulative binomial test. (C) The MAPK Sty1 stress response pathway. All undirected lines represent interactions detected in our interactome. Black arrow represents transcriptional regulation. (D) Sty1-Snr1 interaction validated in *S.pombe* by coimmunoprecipitation (n = 3 blots). (E) Y2H analysis of the ability of Hog1 and Ehd3 to interact and of Sty1 and Snr1 to interact (n = 3 experiments). (F) Sensitivity assays for different deletion strains of *S. cerevisiae* and *S. pombe* under various stress conditions (n = 3 experiments).


A
S. p.
S. c.
Within species
Cross-species
A
B
A′
B′
Intact
Co-evolved
B
Fraction of interactions
P = 0.188
Intact
Co-evolved
C
Skb5
Wis1
Ptc1
Dma1
Sty1
Pyp2
Snr1
Atf1
Hsp16
Conserved
Conserved & Intact
Conserved & Co-evolved
Rewired
D
Myc-Sty1
HA-Snr1
Input lysate
α-Myc IP
α-HA IP
WB: Myc-Sty1
WB: HA-Snr1
WB: γ-Tubulin
1 2 3 4
E
Hog1-Ehd3
Sty1-Snr1
F
Control
YEPD
Osmotic Stress
0.2 mM NaCl
Oxidative Stress
1 mM H2O2
DNA Damage
0.025% MMS
wt
hog1Δ
ehd3Δ
sty1Δ
snr1 Δ
YES
75 mM NaCl
0.75 mM H2O2
0.010% MMS

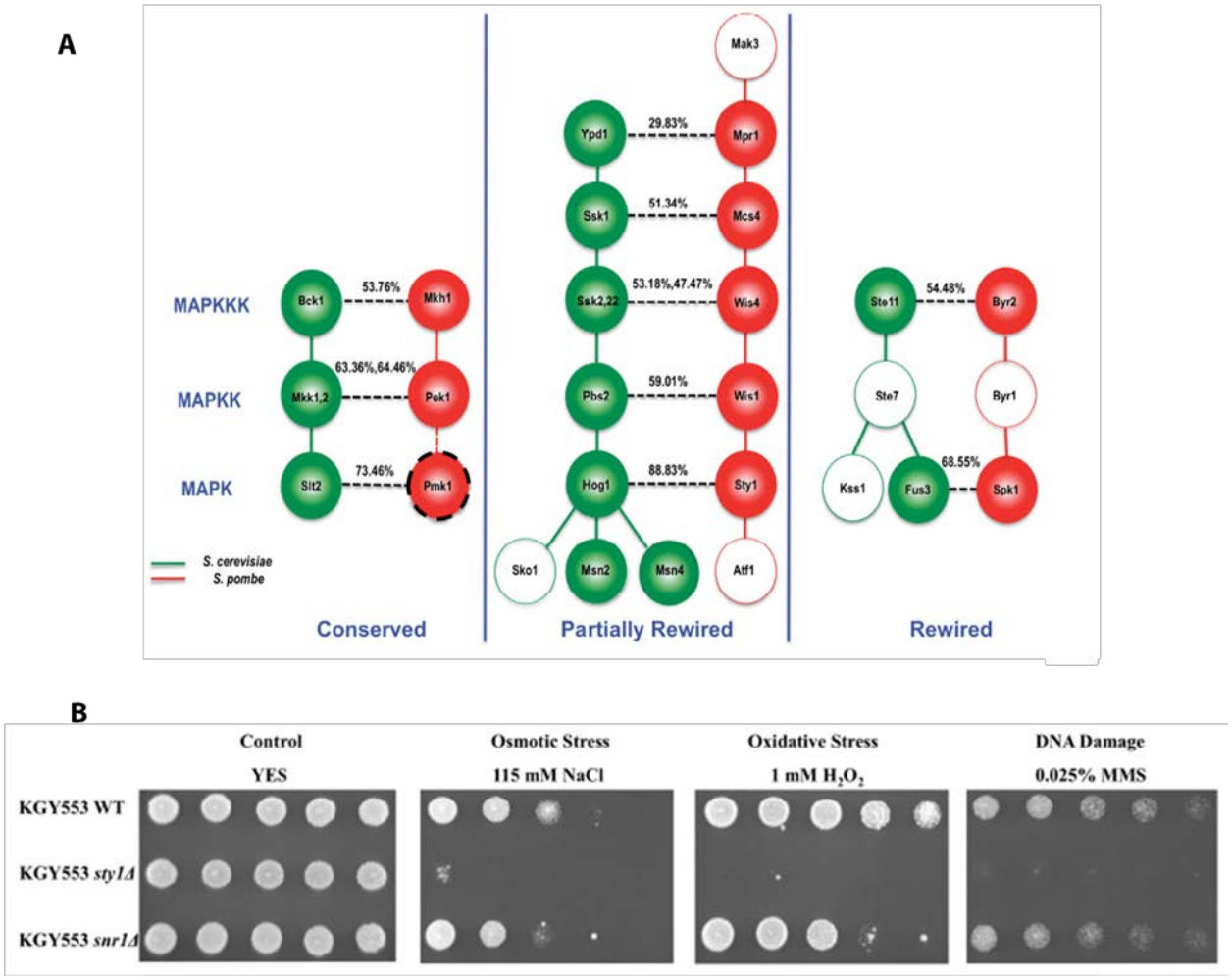


Figure 2.9 Recapitulation of Interactions in the Hog1 and Sty1 Pathways

(A) Filled and empty circles denote proteins with and without corresponding orthologs respectively. Solid lines represent interactions by our HT-Y2H assay, dashed lines represent corresponding orthologous proteins. Values above dashed lines joining orthologous protein pairs correspond to sequence similarity of the two genes. (All proteins except for Pmk1 are in StressNet, hence, this particular protein and the corresponding interaction are denoted by a dashed line.) Pathways are empirically categorized as conserved, partially rewired and rewired based on literature information. (B) Sensitivity assays for deletion strains of *S. pombe* KGY553 (ATCC) under various stress conditions (n = 3 experiments).

not closely related. Furthermore, our results suggest that rewiring of protein interactome networks in related species is likely a major factor for divergence. Surprisingly, we found no significant correlation between the similarity of interaction interfaces and the conservation of corresponding interactions. This demonstrates that conservation of interactions is more complex than previously expected—domains that are not part of the interaction interface also play some indirect role in making the interaction possible. Even if the interface is conserved, the corresponding interaction could still be rewired because of steric hindrance due to altered overall structure or loss of nearby structural scaffolds that make the interaction thermodynamically favorable (Kastritis et al., 2011). We also experimentally explored the evolution of interaction interfaces, and our analysis indicated that interactions traditionally considered "conserved" are equally likely to have intact interfaces as to have coevolved ones that are different from their orthologous counterparts. These results suggest a molecular mechanism by which the interactome network is rewired through evolution: Many proteins have coevolved with their partners to form modified interfaces that can, therefore, accommodate new interactions and functions.

Our results indicated that conserved interactions tended to be stable, and rewired ones were more likely to be transient. Therefore, our finding provides a molecular-level mechanistic explanation for previous studies showing that genetic cross talk between functional modules can differ substantially (Frost et al., 2012; Roguev et al., 2008; Ryan et al., 2012). However, our results also suggest that, overall, proteins tend not to rewire all of their interactions; thus, even if they acquire novel interactions, they still generally conserve at least some of the original functions.

Our results indicate that substantial evolutionary changes, both rewiring and coevolution, of stress response pathways could be a major mechanism by which different organisms adapt to diverse living environments. Conservation of interactions in other pathways might be different from what we observed here. Therefore, similar cross-species interactome mapping and comparative network analyses of more pathways and species will provide a more comprehensive understanding of underlying principles that help shape distinct characteristics of individual organisms through evolution.

2.7 ACKNOWLEDGEMENTS

We thank A. Bretscher for providing budding yeast deletion strains, S. Forsburg for providing fission yeast expression vectors, M. Smolka for experimental advice, and A. Clark and E. Alani for critical reading of our manuscript. J. Das. was supported by the Tata Graduate Fellowship. A. Matsuyama was supported by Grant-in-Aid for Scientific Research. J. A. Pleiss was supported by National Institute of General Medical Sciences (NIGMS) grant GM098634. F. P. Roth was supported by NIH grant HG001715, by the Canada Excellence Research Chairs Program, and by the Canadian Institute for Advanced Research. This work was funded by NIGMS grant R01 GM097358 to H. Yu.

CHAPTER 3

A PROTEOME-WIDE FISSION YEAST INTERACTOME REVEALS NETWORK EVOLUTION PRINCIPLES FROM YEASTS AND HUMAN[1]

3.1 SUMMARY

Here, we present FissionNet, a proteome-wide binary protein interactome for *S. pombe*, comprising 2,278 high-quality interactions, of which ~50% were previously not reported in any species. FissionNet unravels previously unreported interactions implicated in processes such as gene silencing and pre-mRNA splicing. We developed a rigorous network comparison framework that accounts for assay sensitivity and specificity, revealing extensive species-specific network rewiring between fission yeast, budding yeast, and human. Surprisingly, although genes are better conserved between the yeasts, *S. pombe* interactions are significantly better conserved in human than in *S. cerevisiae*. Our framework also reveals that different modes of gene duplication influence the extent to which paralogous proteins are functionally repurposed. Finally, cross-species interactome mapping demonstrates that coevolution of interacting proteins is remarkably prevalent, a result with important implications for studying human disease in model organisms. Overall, FissionNet is a valuable resource for understanding protein functions and their evolution.

---

[1] Reprinted (with permission from the publisher) from Tommy V. Vo[*], Jishnu Das[*], Michael J. Meyer[*], Nicolas A. Cordero, Nurten Akturk, Xiaomu Wei, Benjamin J. Fair, Andrew G. Degatano, Robert Fragoza, Lisa Liu, Akihisa Matsuyama, Michelle Trickey, Sachi Horibata, Andrew Grimson, Hiroyuki Yamano, Minoru Yoshida, Frederick P. Roth, Jeffrey A. Pleiss, Yu Xia, Haiyuan Yu. (2016). A proteome-wide fission yeast interactome reveals network evolution principles from yeasts to human. ***Cell*** 164(1-2):310-23. doi: 10.1016/j.cell.2015.11.037 (Equal contribution is denoted by *).

3.2 CONTRIBUTIONS:

I designed experiments, performed Y2H experiments, produced the DNA microarray heatmaps, performed western blotting and coimmunoprecipitation experiments, performed reverse transcriptase with PCR (RT-PCR) experiments, and all other experiments except for protein complementation assay (PCA) experiments. Jishnu Das performed computational analyses including those pertaining to protein-localization, expression profiles, network conservation, gene duplication, module analyses, and computational support for Y2H experiments. Michael J. Meyer interpreted DNA microarray experiments and performed computational analyses including computational support for Y2H experiments, functional annotation analyses, network conservation analyses, protein interface analyses and predictions, and human diseases. Nicolas A. Cordero performed Y2H experiments. Nurten Akturk performed Y2H and PCA experiments. Xiaomu Wei performed coimmuoprecipitation experiments. Benjamin J. Fair performed DNA microarray experiments and splice-site analyses. Andrew G. Degatano performed Y2H experiments. Robert Fragoza performed Y2H experiments. Lisa Liu performed Y2H experiments. Akihisa Matsuyama generated *S. pombe* entry clones. Michelle Trickey generated *S. pombe* entry clones. Sachi Horibata performed RT-PCR experiments. Andrew Grimson supervised research. Hiroyuki Yamano supervised the generation of *S. pombe* entry clones. Minoru Yoshida supervised the generation of *S. pombe* entry clones. Frederick P. Roth supervised research. Jeffrey A. Pleiss supervised research including DNA microarray experiments and analyses. Yu Xia supervised research including gene duplication analyses. Haiyuan Yu conceived, supervised, and oversaw all aspects of the project, designed experimental and computational analyses. The manuscript was written by me, Jishnu Das, Michael Meyer, and Haiyuan Yu with input from all co-authors.

3.3 INTRODUCTION

Proteins function primarily by physically interacting with other proteins. Gain or loss of these interactions within an organism can modulate protein functions and disease states (Sahni et al., 2015; Wei et al., 2014). The importance of protein interactions to our understanding of fundamental biological processes has spurred the mapping of protein interactome networks for several organisms (Arabidopsis Interactome Mapping Consortium, 2011; Giot et al., 2003; Rolland et al., 2014; Stelzl et al., 2005; Yu et al., 2008). However, the budding yeast *Saccharomyces cerevisiae* remains the only eukaryotic organism for which a high-coverage binary protein interactome has been mapped by systematic interrogation of pairwise combinations of all proteins in triplicate (Yu et al., 2008). Here, we present FissionNet, a high-coverage proteome-wide protein interactome network generated for the fission yeast *Schizosaccharomyces pombe*.

We compared FissionNet with the only other proteome-scale eukaryotic interactomes available (>50% of all protein pairs screened), the interactome networks of *S. cerevisiae* and human. Surprisingly, we find that FissionNet is more similar to the human network than it is to that of *S. cerevisiae*. Furthermore, among interactions involving conserved proteins, there is significant species-specific rewiring that is not completely determined by overall sequence similarity of orthologs. Instead, we identify several other determinants of interaction conservation, including local network constraints and conservation of interacting protein domains. Also, by comparing FissionNet with the proteome-wide interactome of *S. cerevisiae*, we are able to ascertain how gene duplication events influence the process by which paralogs acquire novel functions.

S. pombe is an important model organism for studying fundamental biological processes such as RNA splicing, cell cycle regulation, RNA interference (RNAi), and centromeric maintenance, which are conserved in metazoans but divergent in budding yeast (Wood et al., 2002). We use FissionNet to unveil previously unreported protein associations between gene regulatory factors involved in pre-mRNA splicing and silencing of stress-response genes and at pericentromeric regions, illustrating the value of our network as a proteome-scale resource to understand biological processes.

3.4 MATERIALS AND METHODS

3.4.1 Generation of the binary protein-protein interactome map of *S. pombe*

FissionNet was generated by triplicate independent screening of ~4,900 *S. pombe* ORFs (Figure 3.2A). The network was validated by testing a representative 220 interacting ORF pairs using PCA assays and by determining its functional properties with respect to random pairs and to a literature-curated network.

3.4.2 Conservation of interactions in *S. pombe*, *S. cerevisiae*, and human

We focused only on interactions that can be conserved, *i.e.*, both proteins involved in the interaction have orthologs in the other species We mapped interactions in the reference species to their corresponding ortholog pairs in the other species and tested these pairs using our Y2H assay in a pairwise fashion. Overall, results from these pairwise retests for all three species (a total of ~20,000 individual Y2H experiments) are used to obtain the observed conservation fraction. To accurately estimate the true conservation fraction, we developed a rigorous Bayesian framework that takes into account both the false positive and false negative rates of our Y2H assay, and computes the true conservation fraction from the observed fraction.

3.4.3 Positive and negative reference sets

The PRS and NRS constitute sets of positive and negative controls, respectively. Our PRS comprises 93 *S. pombe* interactions that have been previously reported in 2 or more publications. To construct the NRS we choose 168 random protein pairs that have not been reported to interact in *S. pombe* and whose orthologs have not been reported to interact in any species. In a set of random protein pairs, the expected fraction of interactions is ~10^{-3}-10^{-4} (Riley et al., 2005; Yu et al., 2008), the expected number of interactions in a random set of 168 pairs is <10^{-3}×168 (≈0.2). Since we exclude pairs that are known to interact, the expected number of interactions in our NRS is even lower.

3.4.4 Yeast two-hybrid (Y2H)

Y2H experiments were carried out as previously described by us and other groups (Arabidopsis Interactome Mapping Consortium, 2011; Bandyopadhyay et al., 2010; Boulton et al., 2002; Boxem et al., 2008; Das et al., 2013; Kahle et al., 2011; Lim et al., 2006; Rual et al., 2005; Simonis et al., 2009; Soler-Lopez et al., 2011; Venkatesan et al., 2009; Walhout et al., 2000; Yu et al., 2008; Yu et al., 2011). Briefly, *S. pombe* open reading frames (ORFs) contained in entry vectors (pDONR221) were first cloned into pDEST AD and DB destination vectors using Gateway LR reactions to generate N-terminal ORF fusions. It is important to note that intron-containing ORFs were cloned from cDNA to obtain intronless clones. We will refer to the resultant expression clones as AD-Y and DB-X. After bacterial transformation, miniprepped plasmid DNA of all AD-Y and DB-X expression clones were transformed into Y2H strains *MAT***a** Y8800 and *MAT*α Y8930 (genotype: *leu2-3, 112 trp1-901 his3Δ200 ura3-52 gal4Δ gal80Δ GAL2::ADE2 GAL1::HIS3@LYS2 GAL7::lacZ@MET2 cyh2*^{R}), respectively. Next, we

screened for autoactivators by individually mating each DB-X strain with a *MAT***a** Y8800 strain carrying the empty pDEST AD destination vector. To identify AD autoactivators, we mated each AD-Y strain with a *MAT*α Y8930 strain carrying empty pDEST DB destination vector. Additionally, instead of mating with strains carrying empty pDEST AD or DB vectors, we also tested using strains carrying pDEST AD-eGFP or DB-eGFP. After allowing the yeast to mate on yeast extract peptone dextrose (YEPD) (1% yeast extract, 2% bactopeptone, 2% glucose, 0.45mM adenine sulfate) 2% agar plates at 30°C overnight, yeast were replica plated onto synthetic complete 2% agar plates without leucine and tryptophan (SC+Ade–Leu–Trp+His) and incubated at 30°C overnight to select for diploids with both pDEST AD and DB vector backbones. Finally, diploids were replica plated onto synthetic complete 2% agar plates with 1mM 3-amino-1,2,4-triazole (3AT) and without leucine, tryptophan, and histidine (SC+Ade–Leu–Trp–His+1mM 3AT). Plates incubated at 30°C for 3-5 days. Any AD-Y or DB-Y that grew on SC+Ade–Leu–Trp–His+3AT were scored as autoactivators. The results were consistent using either negative control vectors. We excluded autoactivators from all further screenings.

Thereafter, we performed the first round of testing (called phenotyping I) by mating each unique DB-X with individual mini-pools of ~188 unique AD-Y on YEPD 2% agar plates. We selected for diploids by replica plating onto SC+Ade–Leu–Trp+His. To select for positive interactions, we performed the Y2H screening by replica plating the diploids onto SC+Ade–Leu–Trp–His+3AT and incubating at 30°C for 4 days. We used sterile toothpicks to pick and inoculate all positives into liquid SC+Ade–Leu–Trp+His to keep the yeast in the diploid state.

Next, all yeast colonies picked from phenotyping I were individually subjected to another round of Y2H testing called phenotyping II. Here, all picked colonies were spotted directly onto 2% agar plates of SC+Ade–Leu–Trp–His+3AT and SC–Ade–Leu–Trp+His. Plates were

incubated at 30°C for 4 days. Positives from this round of screening were picked into liquid SC+Ade–Leu–Trp+His to keep the yeast in the diploid state.

Positive yeast picked from phenotyping II were subject to extraction of plasmid DNA by lyses using zymolyase enzyme (Seigakaku Corporation). Cell and enzyme were incubated at 37°C for 45 minutes and then at 95°C for 10 minutes. The identities of DB-X and AD-Y were determined by PCR stitching of the plasmids followed by next-generation sequencing on Illumina MiSeq as described previously (Das et al., 2013; Yu et al., 2011).

Finally, for every AD-Y and DB-X identified by next-generation sequencing (read count ≥ 2), we performed pairwise Y2H testing of each identified pair. This was done by first mating each individual AD-Y with the DB-X putative interacting partner on YEPD plate at 30°C overnight. Then, diploids were selected by replica plating onto SC+Ade–Leu–Trp+His plate and incubating at 30°C overnight. Finally, diploids were selected for interaction-positive cells by replica plating onto SC+Ade–Leu–Trp–His+3AT and SC–Ade–Leu–Trp+His plates and incubating at 30°C for 4-7 days. To identify *de novo* autoactivators (autoactivators that likely arise from accumulation of random mutations during the screening process), we concurrently mated each DB-X with a *MAT***a** Y8800 strain carrying the empty pDEST AD destination vector. Afterwards, we followed the same procedure as during the first autoactivator-detection screen. All identified *de novo* autoactivators were removed from the screens. Thus, at the conclusion of the pairwise Y2H phase, we were able to definitively identify and verify all interacting AD-Y and DB-X while accounting for all *de novo* autoactivators. The entire Y2H pipeline starting from phenotyping I and concluding at pairwise Y2H testing was performed in triplicate (Figure 3.2A).

For Y2H testing of the interactions between Atf1, Cid12, Hrr1, and Rdp1, although the wild-type interactions between these proteins were observed with incubations at 30°C, we also performed longer incubations at 25°C.

3.4.5 Protein Complementation Assay (PCA)

S. pombe ORFs as Gateway entry clones were cloned by Gateway LR reactions into PCA vectors to generate ORF fusions with either of two fragments of yellow fluorescent protein (YFP). Baits were C-terminal fused to the F1 fragment (amino acids 1-158 of YFP) and preys were fused N-terminal with the F2 fragment (amino acids 159-239 of YFP). Each ORF was tested as bait and prey. After bacterial transformation, miniprepped plasmid DNA was prepared on a Tecan Freedom Evo bio-robot. HEK293T cells were grown in DMEM (Life Technologies) supplemented with 10% fetal bovine serum (FBS; Hyclone) at 37°C in a humidified atmosphere containing 5% CO2. Cells were split into 96-well black polystyrene plates (Corning), 24 hours before transfection. Afterwards, 50% confluent cells were transfected transiently with bait and prey plasmids (0.1µg of each plasmid DNA per well for 0.2µg total transfected DNA), using 0.5µl PEI (Polyethylenimine) reagent (1µg/µl, Polysciences Inc.) and 10µl Opti-MEM (Life Technologies). For each 96-well plate, three wells contained HEK293T cells that were untransfected with DNA served as background controls. After 48 hours, YFP fluorescence images were observed and recorded using ImageXpress Micro Widefield automated microscopy (Molecular Devices). YFP fluorescence intensity readings were obtained after 48 hours using a Tecan M1000 plate reader. All intensity readings were first $\log_2$ transformed to normalize signals within each plate. For each plate, pairs were scored positive for interaction if the YFP

fluorescence intensity reading was ≥0.5 (log_2) as compared to the untransfected cells. All pairs scored as positive were verified by fluorescence microscopy images.

3.4.6 *S. pombe* culturing

S. pombe wild-type strains were cultured in yeast extract with supplements (YES) (5g/L yeast extract, 3% glucose, 225mg/L of adenine, histidine, leucine, uracil and lysine hydrochloride). Deletion strains with the *kanMX* cassette were selected for on YES plates with 150mg/L G418 (Calbiochem). Deletion strains with the *natMX* cassette were selected for on YES plates with 100mg/L noursethecin/LEXSY NTC (Mitegen).

S. pombe strains carrying expression plasmids were cultured in minimal media (MM) (1.7g/L yeast nitrogen base without amino acids or ammonium sulfate, 5g/L ammonium sulfate, 2% glucose) with the appropriate supplements. Supplements were added to the media at 225mg/L for adenine, histidine, and leucine. Uracil was added at 50.25mg/L. For protein overexpression from plasmid, strains were grown in Edinburgh minimal media (EMM) (3g/L potassium hydrogen phthalate, 4g/L sodium phosphate dibasic heptahydrate, 5g/L ammonium chloride, 2% glucose, 20mL/L salts, 1mL/L vitamins, 0.1mL/L minerals) with the appropriate supplements. Supplements were added at the above concentrations. Strains harboring pNCH1472 were selected in media lacking uracil supplement. Strains harboring pSGP73 were selected in media lacking leucine supplement. All *S. pombe* strains were grown at 30°C. All *S. pombe* strains used in this study are listed in Table 3.1.

3.4.7 Gene deletion in *S. pombe*

Genes were deleted using a PCR homology-based approach. Briefly, the pFA6a-KanMX6 module was used to replace the entire ORF-of-interest by generating, through stitch

Strain	Genotype		Source
SPY28	h^+	*leu1-32 ura4-D18 imr1R(NCol)::ura4+ oriI ade6-216*	Moazed lab
SPY28 *tas3Δ*	h^+	*leu1-32 ura4-D18 imr1R(NCol)::ura4+ oriI ade6-216 tas3Δ::kanMX6*	This study
SPY28 *chp1Δ*	h^+	*leu1-32 ura4-D18 imr1R(NCol)::ura4+ oriI ade6-216 chp1Δ::kanMX6*	This study
SPY28 *hhp1Δ*	h^+	*leu1-32 ura4-D18 imr1R(NCol)::ura4+ oriI ade6-216 hhp1Δ::kanMX6*	This study
SPY28 *cid12Δ*	h^+	*leu1-32 ura4-D18 imr1R(NCol)::ura4+ oriI ade6-216 cid12Δ::kanMX6*	This study
SPY28 *csl4Δ*	h^+	*leu1-32 ura4-D18 imr1R(NCol)::ura4+ oriI ade6-216 csl4Δ::kanMX6*	This study
SPY28 *atg11Δ*	h^+	*leu1-32 ura4-D18 imr1R(NCol)::ura4+ oriI ade6-216 atg11Δ::kanMX6*	This study
ED666	h^+	*ade6-M210 ura4-D18 leu1-32*	Bioneer *S. pombe* library
srrm1Δ	h^+	*ade6-M210 ura4-D18 leu1-32 srrm1Δ::kanMX6*	Bioneer *S. pombe* library
SPAC30D11.14CΔ	h^+	*ade6-M210 ura4-D18 leu1-32 SPAC30D11.14CΔ::kanMX6*	Bioneer *S. pombe* library
SPAC1952.06CΔ	h^+	*ade6-M210 ura4-D18 leu1-32 SPAC1952.06CΔ::kanMX6*	Bioneer *S. pombe* library
ED668	h^+	*ade6-M216 ura4-D18 leu1-32*	Bioneer *S. pombe* library
ED668 *cid12Δ*	h^+	*ade6-M216 ura4-D18 leu1-32 cid12Δ::kanMX6*	This study
ED668 *dcr1Δ*	h^+	*ade6-M216 ura4-D18 leu1-32 dcr1Δ::natMX6*	This study
ED668 *cid12Δ dcr1Δ*	h^+	*ade6-M216 ura4-D18 leu1-32 cid12Δ::kanMX6 dcr1Δ::natMX6*	This study

Table 3.1 *S. pombe* Strains Used in this Study

PCR, the cassette flanked by ~300bp sequences homologous to 300bp upstream and downstream of the ORF in the genome. The resultant PCR product was purified using Microelute Cycle-Pure kit (Omega) and transformed into *S. pombe* cells using a lithium acetate approach. For generating *dcr1Δ::natMX6* in strain ED668, we obtained a generous gift of *S. pombe* strain PM0414 containing the deletion (from H. Madhani). Thus, we used PCR to amplify the deletion module from strain PM0414 to generate the same deletion in *S. pombe* strain ED668. For generating *cid12Δ::kanMX4,* we used PCR to amplify the deletion module from *cid12Δ* strain in the Bioneer *S. pombe* deletion library. The resultant amplicon was transformed into *S. pombe* strains ED668 and SPY28. All deletion primers used for this study are listed in Table 3.2. The deletion strains of *pwiΔ, SPAC30D11.14CΔ,* and *SPAC1952.06CΔ* were obtained directly from the Bioneer *S. pombe* deletion library.

3.4.8 Identification of Cid12 mutants

In order to select residues integral to the Cid12-Atf1 interaction (*i.e.*, at the interface), but not to Cid12-Hrr1 or Cid12-Rdp1, we used Direct Coupling Analysis (Morcos et al., 2011) to determine evolutionarily correlated residues across interfaces of these interactions in 28 yeast species. Cid12 residues exhibiting the strongest evolutionary couplings with Atf1 residues were considered likely to facilitate the Cid12-Atf1 interaction. In order to increase the chances of selecting Cid12 residues that are not at the interaction interface of other Cid12 interactions, we did not consider any Cid12 residues with strong evolutionary couplings with Hrr1 or Rdp1. Once Cid12 residues were chosen, we introduced amino acid mutations designed to strongly alter the hydrophobicity of the wild-type amino acid.

Primer	Sequence (5' → 3')	Purpose
Tas3-UP-fwd	CTTTAATGAAGAGGTCAAGG	Deletion of *tas3*
Tas3-UP-rev	TTAATTAACCCGGGGATCCGTAAAATCGTAGCTCAAATAG	Deletion of *tas3*
Tas3-pFA-Kan-fwd	CTATTTGAGCTACGATTTTACGGATCCCCGGGTTAATTAA	Deletion of *tas3*
Tas3-pFA-Kan-rev	CTTGAGTCAGTAAACTACTTGAATTCGAGCTCGTTTAAAC	Deletion of *tas3*
Tas3-DN-fwd	GTTTAAACGAGCTCGAATTCAAGTAGTTTACTGACTCAAG	Deletion of *tas3*
Tas3-DN-fwd	GTGTACAACAATCATTTAGG	Deletion of *tas3*
Hhp1-UP-fwd	AGCATCACGCTCCCATTCTA	Deletion of *hhp1*
Hhp1-UP-rev	TTAATTAACCCGGGGATCCGGCACTCGCTTTTTCTCTCCA	Deletion of *hhp1*
Hhp1-pFA-Kan-fwd	TGGAGAGAAAAAGCGAGTGCCGGATCCCCGGGTTAATTAA	Deletion of *hhp1*
Hhp1-pFA-Kan-rev	CGTCTGGAGGATAAAAGGCAGAATTCGAGCTCGTTTAAAC	Deletion of *hhp1*
Hhp1-DN-fwd	GTTTAAACGAGCTCGAATTCTGCCTTTTATCCTCCAGACG	Deletion of *hhp1*
Hhp1-DN-rev	TTTTCTGTTCCTGGCTCGAA	Deletion of *hhp1*
Cid12-deletion-fwd	CAATGAGAGATGAGGGAG	Deletion of *cid12*
Cid12-deletion-rev	GAACAGATTTTGTTAGCAC	Deletion of *cid12*
Dcr1-deletion-fwd	CTACCACCTACTTTTCTC	Deletion of *dcr1*
Dcr1-deletion-rev	CAACATCTATGAACGATATC	Deletion of *dcr1*
Chp1-UP-fwd	CTAAAAGCTTCATTACTGG	Deletion of *chp1*
Chp1-UP-rev	TTAATTAACCCGGGGATCCGTTCAAGGTAGCGGGTT	Deletion of *chp1*
Chp1-pFA-Kan-fwd	AACCCGCTACCTTGAACGGATCCCCGGGTTAATTAA	Deletion of *chp1*
Chp1-pFA-Kan-rev	CCAAAATAAATAAAACGAGAATTCGAGCTCGTTTAAAC	Deletion of *chp1*
Chp1-DN-fwd	GTTTAAACGAGCTCGAATTCTCGTTTTATTTATTTTGG	Deletion of *chp1*
Chp1-DN-rev	TTGTTTAGTTTAAATACAC	Deletion of *chp1*

Table 3.2 Primers Used for this Study

Primer	Sequence (5' → 3')	Purpose
Csl4-UP-fwd	TTTTGGCATCGTTTTGTGAA	Deletion of *csl4*
Csl4-UP-rev	TTAATTAACCCGGGGATCCGAAAGGAATCAAGTCCGTTTTG	Deletion of *csl4*
Csl4-pFA-Kan-fwd	CAAAACGGACTTGATTCCTTTCGGATCCCCGGGTTAATTAA	Deletion of *csl4*
Csl4-pFA-Kan-rev	CAAAGCAATTTTATCCCCTCAGAATTCGAGCTCGTTTAAAC	Deletion of *csl4*
Csl4-DN-fwd	GTTTAAACGAGCTCGAATTCTGAGGGGATAAAATTGCTTTG	Deletion of *csl4*
Csl4-DN-rev	TTCAGAGCTCCTGTCCGTTC	Deletion of *csl4*
Atg11-UP-fwd	GAGGGTATACGGCTACATATGACA	Deletion of *atg11*
Atg11-UP-rev	TTAATTAACCCGGGGATCCGGCGATTCCCCAACATTAAAC	Deletion of *atg11*
Atg11-pFA-Kan-fwd	GTTTAATGTTGGGGAATCGCCGGATCCCCGGGTTAATTAA	Deletion of *atg11*
Atg11-pFA-Kan-rev	TCAGTCCCACAAATACAAGCAGAATTCGAGCTCGTTTAAAC	Deletion of *atg11*
Atg11-DN-fwd	GTTTAAACGAGCTCGAATTCTGCTTGTATTTGTGGGACTGA	Deletion of *atg11*
Atg11-DN-rev	GCGTTGTTCGGTGTGATAAA	Deletion of *atg11*
Cid12-mut-D260V-fwd	CTATCTATTGAAGATCCAATTGTCAGAAATAACGACATTGGAAAG	Generate cid12^{D260V} mutant
Cid12-mut-D260V-rev	CTTTCCAATGTCGTTATTTCTGACAATTGGATCTTCAATAGATAG	Generate cid12^{D260V} mutant
Cid12-mut-K213I -fwd	CAATTCGAGCTTTACTACAGATATTCTTCTATTTTTGGGGAG	Generate cid12^{K213I} mutant
Cid12-mut-K213I -rev	CTCCCCAAAAATAGAAGAATATCTGTAGTAAAGCTCGAATTG	Generate cid12^{K213I} mutant
Tas3-pNCH1472-fwd	AAGGAAAAAAGCGGCCGCATGGAGAAAGGAATC	Generate tas3-myc
Tas3-pNCH1472-rev	GGTGTCGACTTTTTCGTTTTTATGCTC	Generate tas3-myc
Hhp1-pSGP73-fwd	AAGGAAAAAAGCGGCCGCGCTTTGGACCTCCGGATT	Generate HA-hhp1
Hhp1-pSGP73-rev	GGAAGATCTCTAATTAGGTCTGTTGATATATTG	Generate HA-hhp1

(Table 3.2 continued)

Primer	Sequence (5' → 3')	Purpose
Atf1-pNCH1472-fwd	AAGGAAAAAAGCGGCCGCATGTCCCCGTCTCC	Generate atf1-myc
Atf1-pNCH1472-rev	GGTGTCGACGTACCCTAAATTGATTCTTTG	Generate atf1-myc
Cid12-pSGP73-fwd	AAGGAAAAAAGCGGCCGCGGTAAAGTCCTGTTAGAG	Generate HA-cid12
Cid12-pSGP73-rev	GGAAGATCTTTATCCGCCAGCTTG	Generate HA-cid12
cen dh-fwd	GAAAACACATCGTTGTCTTCAGAG	ChIP
cen dh-rev	CGTCTTGTAGCTGCATGTGAA	ChIP
fbp1-fwd	GGTTGCTGCTGGCTATACTATG	ChIP
fbp1-rev	TGGATAAGCAAACAACCCACC	ChIP

(Table 3.2 continued)

3.4.9 Site-directed mutagenesis

Specific point mutations were introduced into *S. pombe* ORF entry clones using site-directed mutagenesis as previously described (Wei et al., 2014). Briefly, primers were designed to introduce the point mutations and to generate PCR products that could self-circularize. Template DNA used in the PCR reactions was digested away by incubation with *Dpn*I overnight at 37°C. The digested products were transformed into DH5α bacterial cells. Single-colony transformants were individually picked and Sanger sequenced to verify clones containing only the intended mutations. All mutagenesis primers used for this study are listed in Table 3.2.

3.4.10 Western blotting and protein coimmunoprecipitation

Strains were first grown in EMM with the appropriate supplements at 30°C for 18 hours to induce protein expression. Then, cells were collected by centrifugation at 700g for 5 min. Cells were lysed using cold glass beads in lysis buffer (50mM tris-HCl pH 7.5, 0.2% tergitol, 150mM NaCl, 5mM EDTA) containing protease inhibitor cocktail (Roche) and 5mM PMSF. Whole-cell extracts (WCE) were collected and normalized using Bradford assay or Nanodrop.

Immunoprecipitations were performed by incubating WCE with EZview red anti-HA affinity gel (Sigma-Aldrich) overnight at 4°C. Following extensive washing using cold lysis buffer, the anti-HA affinity gels were resuspended into Laemlli SDS-PAGE buffer and denatured by boiling for 5 minutes. For controls, the same amount of WCE input were also resuspended in the same buffer and denatured. Proteins were resolved in 10% SDS-PAGE gel electrophoresis and transferred onto PVDF membrane (GE Amersham). Protein detection was determined using either primary rabbit anti-myc (Santa Cruz) or mouse anti-HA (12CA5, Roche) antibody and secondary HRP anti-rabbit IgG (Cell Technologies) or HRP anti-mouse IgG (Abcam) antibody.

Signal development was performed using ECL Prime Western Blot Detection kit (GE Amersham). All washing and blocking steps were performed using 5% non-fat dry milk and TBS with 0.1% tween-20.

Western blots to detect protein expression were performed using 10% SDS-PAGE electrophoresis and transfer to PVDF membrane (GE Amersham). Protein detection was determined using primary rabbit anti-myc (Santa Cruz), mouse anti-HA (12CA5, Roche), or rabbit anti-actin (MP Biomedicals) antibody and secondary HRP anti-rabbit IgG (Cell Technologies) or HRP anti-mouse IgG (Abcam) antibody. All washing and blocking steps were performed using 5% non-fat dry milk and TBS with 0.1% tween-20.

3.4.11 Centromeric silencing assays

S. pombe strains used for these assays contained the *ura4*$^{+}$ reporter gene within the endogenous centromeric *imr1R* region. Wild-type cells do not express the reporter gene from this locus due to normal centromeric silencing and, thus, the cells are insensitive to the drug 5-fluoro-orotic acid (5-FOA) (Zymo Research). Cells were grown in YES media to log phase (OD_{600}=0.7-1.0) at 30°C. Strains were then normalized to OD_{600}=0.1 in 1mL YES and, then, four-fold serial dilutions were made. Lastly, cells were spotted onto non-selective (N/S) plates, -Ura plates, and onto 0.1% 5-FOA plates. All plates were incubated at 30°C for 3-4 days.

S. pombe strains carrying either empty pSGP73 vector or pSGP73 vector with wild-type or mutant *cid12* were grown in MM+Ade+Ura–Leu to log phase at 30°C. Afterwards, they were normalized to OD_{600}=0.1 in 1mL MM+Ade+Ura–Leu and, then, four-fold serial dilutions were made. Finally, cells were spotted onto N/S plates (EMM supplemented with 225mg/L adenine

and 50.25mg/L uracil) and onto 5-FOA plates (EMM supplemented with 225mg/L adenine, 50.25mg/L uracil, and 5-FOA). Plates were incubated at 30°C for 2-3 days.

3.4.12 Reverse transcription PCR (RT-PCR)

RNA isolation procedures were performed as described in (Inada and Pleiss, 2010). Strains were first grown until log phase (OD_{600}=0.7-1.0) at 30°C. Then, 10mL cells were collected by centrifugation at 700g for 5 minutes. Cells were lysed in 2mL acid-phenol chloroform (Amresco) and AES buffer (50mM sodium acetate pH 5.3, 10mM EDTA, 1% SDS) at 65°C with short 5 – 10 second vortex pulses every 1 minute for 7 minutes. Using Phase Lock Gel Heavy tubes (5 Prime), WCE was extracted and subject to sequential RNA purification from contaminant proteins and DNA by addition of phenol:chloroform:IAA (Ambion) and chloroform (Alfa Aesar). Each 2mL of purified RNA was precipitated in 2.2mL isopropanol with 200µL 3M sodium acetate pH 5.3. RNA pellets eventually dissolved into DEPC-treated MilliQ water.

Synthesis of cDNA from 1µg purified RNA was performed using a high capacity cDNA reverse transcription kit (Applied Biosystems) according to manufacturer protocol. Thereafter, cDNA was diluted 5-fold. For semi-quantitative RT-PCR, 1µL of diluted cDNA was subject to 25-cycles of PCR and resolved in 1.5% agarose gels. Primers used were *hsp16* (forward 5’ AAAGCACCGAGGGTAAC-CAA -3’ and reverse 5’- TGGTACGAGAGAATGAGCCAAA -3’), *hsp104* (forward 5’-CGTGAATCT-CAGCCCGAAGT -3’ and reverse 5’- TCAACGCGGAGTTGTCGAA -3’), and *act1* (forward 5’- TCCTCATGCTATCATGCGTCTT -3’ and reverse 5’- CCACGCTCCATGAGAATCTTC -3’).

Quantitative RT-PCR was performed using Power SYBR green PCR master mix (Applied Biosystems). SYBR green signal was detected by StepOne Real-Time PCR System

machine (Applied Biosystems) and quantification was performed by StepOne software v2.0. *Otr dh* primers used were (forward 5'- TTCTGAATAATTGGGATCGC -3' and reverse 5'- TGCTGTCATACTACACTGCA 3'). The same *act1* primers above were used for quantitative RT-PCR.

3.4.13 Chromatin immunoprecipitation (ChIP)

Cells were grown at 30°C and log-phase cultures were harvested. Afterwards, cells were fixed in 1% formaldehyde for 20 min at room temperature followed by glycine-quenching for an additional 5 min. Cells were lysed using cold glass beads and chromatin was sheared by sonication using a Branson Digital (Emerson) Sonifier at 30% amplitude for 7 min of 10 sec ON and 1 min OFF. As input control, 5% of chromatin was used. For H3K9me2 ChIP, 3µL (3µg) of mouse anti-H3K9me2 (Abcam) was used per ChIP. Chromatin and antibody were incubated 4°C for 2 hours with rotation. Then, 25µL of washed protein A/G-agarose resin (Santa Cruz) was added and incubated for an additional 1 hour with rotation. After washing and elution from the resin, crosslink reversal was performed by incubation at 65°C overnight. DNA was precipitated by phenol/chloroform extraction. Multiplex PCR was performed using primers for *cen dh* and *fbp1* that are listed in Table 3.2. PCR products were resolved on 1.5% agarose gel with ethidium bromide.

3.4.14 Splicing-specific DNA microarray (Sample preparation and microarray design)

S. pombe cells and RNA were processed for DNA microarray experiments as described previously (Inada and Pleiss, 2010). Briefly, wild-type or deletion strains were grown in YES until log phase, after which they were harvested and lysed for total RNA collection. To test the

effect of a given gene deletion on global intronic splicing, total RNA from each deletion strain and the corresponding wild type strain were converted to cDNA and then labeled with either Cy3 or Cy5 dye. The labeled cDNA were allowed to competitively bind to a DNA microarray slide. Microarray images were captured using an Axon Instruments GenePix 4000B scanner and analyzed using Axon Instruments GenePix Pro software. Spot ratio data were preprocessed by $\log_2$ transformation. The DNA microarray (Agilent) was designed to contain three types of oligo probes per *S. pombe* mRNA: (1) a single probe to identify each specific mRNA (both pre-mRNA and spliced mRNA); this probe recognizes the longest exon of the mRNA, (2) a probe corresponding to each known intron present in a pre-mRNA species to measure the level of unspliced pre-mRNA, and finally (3) a probe recognizing every possible exon-intron junction to measure the levels of mature mRNA.

3.4.15 Splicing-specific DNA microarray (Calculation of 5' splice site log-scores)

Introns with ≥0.5 ($\log_2$) fold enrichment in the mutant relative to wild-type were categorized as "affected". Otherwise, introns were categorized as "unaffected". The log-odds score of each intron's annotated 5' splice-site (Wood et al., 2012) was determined using a weight matrix model (Lim and Burge, 2001). The set of affected log-odds scores was compared to the set of unaffected with a *U* test.

3.4.16 Detection rates of the positive (PRS) and negative (NRS) reference sets

Of the 168 NRS pairs, 78 are between proteins with different sub-cellular localizations and 90 have the same sub-cellular localization (Matsuyama et al., 2006). However, there is no significant difference in the fraction of random pairs detected by either Y2H (0/78 and 0/90 for

the two sets respectively, *P*=0.92 using a *Z* test) or PCA (8/78 and 9/90 for the 2 sets respectively, *P*=0.96 using a *Z* test).

To examine if there are any species-specific biases of our Y2H assay, we computed the fractions of PRS and PRS-nonY2H (subset of PRS interactions that have been detected using an assay other than Y2H) interactions in the three different species that are recapitulated by our Y2H assay. We find that there is no significant difference between the detection rates across species (Figure 3.6C; *P*>0.35 for all pairs, *Z* test). Furthermore, we find that there is no significant difference in interaction density (*i.e.*, number of interactions detected divided by total number of protein pairs screened) for FissionNet and previously reported Y2H interactomes in *S. cerevisiae* (Yu et al., 2008) and human (Rolland et al., 2014) (Figure 3.6D; all interaction densities differ by <2 fold). These results confirm that our Y2H assay has no species-specific detection biases.

3.4.17 Calculating the coexpression of genes

To measure the coexpression of transcripts corresponding to proteins involved in FissionNet interactions, we calculated the Pearson Correlation Coefficient (*PCC*) between their expression profiles: expression values measured at different time-points in the cell cycle (Rustici et al., 2004). We also calculated the *PCC* between expression profiles of transcripts corresponding to proteins involved in high-quality *S. pombe* interactions from literature curation. Finally, we defined two different sets of random pairs: (1) all random pairs, (2) random pairs by permuting edges between proteins in the network. We first compared the different distributions using a KS test. Next, we calculated the fractions of significantly co-expressed interactions, as well as the fraction of significantly co-expressed random pairs. We defined significant

coexpression as *PCC* ≥ a threshold value. To ensure that our conclusions are robust to the choice of this threshold, we used three different thresholds: 0.2, 0.4 and 0.5 (Figures 3.1C-H). When comparing the fractions of interactions or pairs that are significantly co-expressed, *P*-values were calculated using a *Z* test.

3.4.18 Other functional properties of FissionNet

For other calculations, since small-scale studies could focus on proteins with more complete annotations in GO, we restricted our analyses to a set of proteins found in both high-quality literature-curated *S. pombe* interactions (Das and Yu, 2012) and interactions in FissionNet. We then defined 3 sets of protein pairs such that both proteins are from the previously defined set: (1) high-quality *S. pombe* interactions from literature curation, (2) *S. pombe* interactions from FissionNet, and (3) all pairs of proteins for which the two proteins have never been reported to interact. We performed the following calculations on these 3 sets:

Calculating functional similarity

We calculated functional similarity using a total ancestry method that computes all pairwise functional similarities in a set of proteins by determining for each given pair of proteins, the number of other protein pairs sharing the same set of parent GO terms (Yu et al., 2007). In this framework, a pair of proteins that are very dissimilar will share their GO ancestry with a large number of other protein pairs. Conversely, a pair of proteins that are very similar will share their GO ancestry with only a few or none of the other pairwise combinations of proteins in the same set. Each similarity score for a pair of proteins was computed as a percentile ranking of their total ancestry score among all such scores calculated for all pairwise combinations of

proteins in the set. We considered the top 1% of protein pairs in this ranking to be functionally similar. *P*-values were calculated using a *Z* test.

Calculating co-localization

To calculate the co-localization of proteins involved in FissionNet interactions, we calculated the fraction of protein pairs that have the same sub-cellular localization (Matsuyama et al., 2006). *P*-values were calculated using a *Z* test.

3.4.19 Conservation of genes

To analyze the extent to which genes are conserved, we calculated the fraction of genes in the reference species *i* that also have orthologs in the other species *j*:

$$Gene\_cons_{ij} = \frac{G_i^{j'}}{G_i}$$

where G_i denotes the total number of genes in species *i* and $G_i^{j'}$ the number of genes in species *i* that have corresponding orthologs in species *j*. Using ortholog annotations from PomBase and the Saccharomyces Genome Database, we computed the extent of gene conservation between different species pairs for all coding genes (Figure 3.6A) (Cherry et al., 2012; McDowall et al., 2015). We also used orthologs from InParanoid to compute the extent of gene conservation between different species pairs for all coding genes (Figure 3.6B) (Sonnhammer and Ostlund, 2015). We observe the same gene conservation trends regardless of which database is used for determining orthology, confirming the robustness of our result.

3.4.20 Estimating true interaction conservation fractions

To calculate the extent to which interactions are conserved, we focused only on those interactions that can be conserved, *i.e.*, both proteins involved in the interaction have orthologs in the other species. For each pair of organisms, we used both organisms as the reference (six comparisons for 3 species). We mapped interactions in the reference species to their corresponding ortholog pairs in the other species and tested these pairs using our Y2H assay in a pairwise fashion. We performed pairwise retests because we have shown earlier that not all interactions detected by Y2H in a pairwise fashion will be detected in a high-throughput screen where individual baits are tested against minipools of ~188 preys (Yu et al., 2008). Overall, results from these pairwise retests for all three species (a total of ~20,000 individual Y2H experiments) are used to obtain the observed conservation fraction. To accurately estimate the true conservation fraction, we used a rigorous Bayesian framework that takes into account both the false positive and false negative rates of our Y2H assay, and computes the true conservation fraction from the observed fraction.

Using the law of total probability, we can write:

$$P(D \mid I') = P(D \mid I,\ I') \times P(I \mid I') + P(D \mid \bar{I},\ I') \times P(\bar{I} \mid I') \quad (1)$$

Here, I' denotes the event that an interaction occurs in the reference species, I the event that the interaction occurs in another species, $\bar{I}$ the event that the interaction does not occur in the other species and D the event that it is detected in the other species using our Y2H pipeline. The observed conservation rate is $P(D|\ I')$. The true conservation rate is $P(I|\ I')$. As an interaction in the reference species can only be either conserved or rewired in the other species:

$$P(I \mid I') + P(\bar{I} \mid I') = 1 \quad (2)$$

Finally, we can assume conditional independence between D and I' given I. In other words, given that an interaction occurs in the other species, whether it is Y2H detectable in that species and whether its ortholog pair interacts in the reference species are independent of each other. Using this:

$$P(D \mid I, I') = \frac{P(D,I,I')}{P(I,I')} = \frac{P(D, I' \mid I) \times P(I)}{P(I,I')} = P(D \mid I) \times \left(\frac{P(I' \mid I) \times P(I)}{P(I,I')} \right) = P(D \mid I) \quad (3)$$

Using similar arguments,

$$P(D \mid \bar{I}, I') = P(D \mid \bar{I}) \quad (4)$$

Substituting equations (2), (3) and (4) in equation (1), we obtain:

$$P(I \mid I') = \frac{P(D \mid I') - P(D \mid \bar{I})}{P(D \mid I) - P(D \mid \bar{I})} \quad (5)$$

$P(D| I')$ is estimated using the fraction of interactions in the reference species that are detected by Y2H to interact in the other species (f_d). $P(D|I)$ is estimated using the fraction of a set of true interactions (PRS) that we can detect using our Y2H assay (f_{prs}). Finally, $P(D| \bar{I})$ is estimated using the fraction of a set of random pairs that are unlikely to interact (NRS) that we can detect using our Y2H assay (f_{nrs}). So, for any species pairs:

$$P(I \mid I') = \frac{f_d - f_{nrs}}{f_{prs} - f_{nrs}} (6)$$

We can estimate the error using the delta method:

$$SE_{P(I|I')} = \sqrt{\frac{\left(f_{prs} - f_{nrs}\right)^2 \times \left(SE_{f_d}\right)^2 + \left(f_{prs} - f_d\right)^2 \times \left(SE_{f_{nrs}}\right)^2 + \left(f_d - f_{nrs}\right)^2 \times \left(SE_{f_{prs}}\right)^2}{\left(f_{prs} - f_{nrs}\right)^4}} \quad (7)$$

3.4.21 Interaction conservation using assays other than Y2H

We examined the observed conservation as detected by individual assays rather than using overall interactome networks from the literature as these are derived from assays with varied and unknown false positive and false negative rates. However, for a single assay with

unknown false positive and false negative rates, while we will be unable to calculate the true underlying conservation fraction, we can still compute the observed conservation fraction. We first calculated the fraction of FissionNet interactions whose corresponding *S. cerevisiae* and human ortholog pairs have been shown to interact in co-crystal structures (Das and Yu, 2012). We find that that fission yeast interactions are better conserved in human than in budding yeast (Figure 3.6G; >2 fold difference in observed conservation, $P<10^{-3}$). Next, we calculated the fraction of FissionNet interactions whose corresponding *S. cerevisiae* and human ortholog pairs have been detected as interacting by proteome-scale affinity purification/mass spectrometry experiments (Gavin et al., 2006; Huttlin et al., 2015; Krogan et al., 2006). Here, we also find that fission yeast interactions are better conserved in human than in budding yeast (Figures 3.6H and 3.6I; >1.5 fold difference in observed conservation, $P<10^{-3}$ in both cases).

3.4.22 Identifying proteins conserved in eukaryotes

To identify proteins that are conserved across eukaryotes, we used clusters of conserved eukaryotic orthologous groups of genes (KOGs) as defined by Koonin *et al.* (Koonin et al., 2004). These conserved KOGs often comprise genes essential for survival and could be considered to approximate "a minimal set of essential eukaryotic genes" (Koonin et al., 2004). Each KOG consists of orthologous genes in up to 7 representative eukaryotic species studied by the authors. We defined proteins conserved in eukaryotes as those proteins from these KOGs that are conserved in ≥5 species.

3.4.23 Interaction conservation in different biological processes

We used the Gene Ontology (GO) (Ashburner et al., 2000) to categorize interactions based on the annotations of the proteins involved. We computed interaction conservation in GO Slim Biological Process (BP) categories, a set of 70 terms representative of diverse biological processes not specific to any one organism. For all analyses, we considered only genes annotated with experimental evidence codes (Ashburner et al., 2000). We considered an interaction to be within a category if either of its interacting proteins is annotated in that category or one of its children.

3.4.24 Sequence conservation of proteins and interactions

To determine the sequence conservation between two proteins, alignments were produced using the pairwise2.align.global function of the BioPython Python module, an implementation of the Needleman-Wunsch global alignment algorithm (Needleman and Wunsch, 1970). We used the BLOSUM62 scoring matrix, a gap-open penalty of -10 and a gap-extend penalty of -0.5. Two amino acids are considered similar if the BLOSUM62 score associated with a substitution between the two residues is >0. Unless otherwise specified, sequence similarity is measured with sequences of *S. pombe* proteins serving as the reference. Sequence similarity between an *S. pombe* protein and an ortholog in another species is measured as the fraction of *S. pombe* residues similar to their aligned residues in either *S. cerevisiae* or human. To calculate the sequence similarity of pairs of proteins with orthologous pairs, the individual sequence similarities of each protein with their orthologs are averaged. *P*-values were calculated using a *Z* test.

3.4.25 Interface domain conservation based on co-crystal structures

We compiled a set of co-crystal structures from the PDB representing human protein-protein interactions. For each structure, we calculated interface residues using NACCESS to determine surface residues whose solvent accessible surface area was altered by $\geq 1 Å^2$ between bound and unbound states (Hubbard, 1996). To determine interface residues of protein interactions, we took the union of interface residues determined from each representative PDB chain pair for which at least 5 interface residues were calculated in each chain. In the human interactions, we identified Pfam domains at the interaction interface as those domains containing at least 5 interface residues. All domains not meeting this criterion are considered 'Other' as we don't know if they facilitate the interaction or not. We then aligned the full human protein sequences in each interaction to their orthologs in *S. pombe* and *S. cerevisiae* using the alignment method mentioned previously. Here, we used the human sequences as the reference and only calculated sequence similarity within the portions of the alignment in the human domain regions. *P*-values were calculated using a *U* test.

3.4.26 ClusterOne

We performed clustering with ClusterONE (Nepusz et al., 2012). ClusterONE finds overlapping functional modules and is specifically tuned for clustering biological networks. We used ClusterONE with parameters *s*=3 (minimum cluster size) and *d*=0.5 (minimum cluster density) and found 193 clusters in our network. Since proteins can belong to multiple clusters, we defined an intra-cluster interaction as any interaction for which there is a cluster that contains both proteins and an inter-cluster interaction as any interaction for which both proteins belong to clusters, but there is no cluster that contains both proteins. Intra-cluster and inter-cluster

conservations were calculated using the fraction of interactions within and across clusters that are detected as conserved using our Y2H assay, transformed via the Bayesian framework described above to obtain the true conservation fractions (Figure 3.7E). *P*-values were calculated using a *Z* test.

3.4.27 Affinity propagation clustering

We also used affinity propagation clustering (APC) to generate clusters from FissionNet (Frey and Dueck, 2007). APC relies only on the topological properties of the network to generate clusters. The algorithm requires a pairwise similarity measure as input. This was defined as follows:

$$Sim_{i,j} = (1 + Diam_{FN}) - Dist_{i,j}$$

Here $Dist_{i,j}$ is the graph distance between nodes *i* and *j*. $Diam_{FN}$ is the graph diameter of FissionNet, *i.e.*, the maximum distance between any two nodes. Graph distance was set to 1 + $Diam_{FN}$ for node pairs that are not connected. Thus, $Sim_{i,j}$ represents the normalized similarity between two nodes. It will take the highest value (equal to $Diam_{FN}$) for nodes that are directly connected and the lowest value (0) for nodes that are not connected at all.

Observed intra-cluster and inter-cluster conservations were obtained by calculating the fraction of interactions within and across clusters that are detected as conserved using our Y2H assay (Figure 3.8A). *P*-values were calculated using a *Z* test.

3.4.28 Gene Ontology

We used Gene Ontology (GO) (Ashburner et al., 2000) and GO Slim annotations from PomBase (McDowall et al., 2015) to cluster FissionNet based on known biological processes. Intra-process and inter-process conservations were calculated using the fraction of interactions

within and across processes that are detected as conserved using our Y2H assay (Figure 3.8B; GO Slim), and transformed via the Bayesian framework described above to obtain the true conservation fractions (Figure 3.7F; all GO). *P*-values were calculated using a *Z* test.

3.4.29 Distribution of intact and coevolved interactions across species

We computed the log-odds ratios for 3 scenarios: an interaction is intact in both species pairs (*S. pombe*-*S. cerevisiae* and *S. pombe*-human), an interaction is coevolved in both species pairs, an interaction is intact in one species pair but coevolved in the other:

$$LOR = \log\left(\frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}\right)$$

where, p_1 is the observed fraction of interactions in each category and p_2 the expected fraction of interactions in each category. The expected fraction is calculated assuming independence between the events of being intact/coevolved in each species pair. Standard error was calculated using the delta method:

$$SE_{LOR} = \sqrt{\left(\frac{SE_{p1}{}^2}{p_1{}^2 \times (1-p_1)^2} + \frac{SE_{p2}{}^2}{p_2{}^2 \times (1-p_2)^2}\right)}$$

P-values were calculated using a *Z* test.

3.4.30 Sub-functionalization and neo-functionalization

We obtained a set of 3,853 fission yeast paralog pairs and 6,846 budding yeast paralog pairs from Ensembl Biomart (Kinsella et al., 2011). For budding yeast, WGD paralogs were defined based on annotations from Kellis *et al.* (Kellis et al., 2004), and the rest were considered to be SSD paralogs.

To measure the extent of sub-functionalization, we calculated the fraction of interactions that are conserved but not shared among paralog pairs. We normalized this by the fraction of conserved but not shared interactions among all pairs of proteins that do not have a paralog. The fraction of conserved and not shared interactions is equal to 1 – the fraction of conserved and shared interactions. To calculate the fraction of conserved and shared interactions, we first constructed a set of high-quality interactions from the literature that are conserved. If we use the literature to ascertain how many of these interactions are shared, the fraction will be inaccurate as literature-curated interactomes are incomplete and suffer from detection rate biases. To circumvent this, we first calculated the fraction of conserved interactions that were detected as shared using our Y2H assay. We then used our previously developed framework to calculate the actual number of shared interactions:

$$f_{shared} = \frac{f_{obs} \times precision}{completeness \times \text{assay\_sensitivity} \times \text{sampling\_sensitivity}}$$

where f_{obs} is the detected fraction of shared pairs using our Y2H assay and f_{shared} is the actual fraction of shared pairs (Yu et al., 2008). Precision, completeness, assay-sensitivity, and sampling-sensitivity for FissionNet are calculated as previously described (Yu et al., 2008). For the CCSB-YI1 network, they have been previously reported (Yu et al., 2008). With the calculated fractions of conserved and not shared interactions, we computed the following log odds ratio for *S. pombe* and *S. cerevisiae*:

$$LOR = \log\left(\frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}\right)$$

where p_1 is the fraction of interactions that are conserved with its ortholog but not shared among paralog pairs and p_2 by the fraction of conserved but not shared interactions among all pairs of

proteins that do not have a paralog. Standard error was calculated using the delta method as described earlier. *P*-values were calculated using a *Z* test.

To measure the extent of neo-functionalization, we calculated the log odds ratio of the fractions of rewired interactions involving proteins that have and do not have paralogs. We computed the same log odds ratio for *S. pombe* and *S. cerevisiae*:

$$LOR = \log\left(\frac{\frac{p_3}{1-p_3}}{\frac{p_4}{1-p_4}}\right)$$

where, p_3 is the fraction of rewired interactions involving proteins where at least one has a paralog and p_4 the fraction of rewired interactions between proteins that do not have paralogs. The fraction of rewired interactions is defined as 1 – the fraction of conserved interactions. Since we are able to calculate the true fraction of conserved interactions using a Bayesian framework that accounts for assay false positive and negative rates (please refer to '**Conservation of interactions in *S. pombe*, *S. cerevisiae*, and human**'), the fraction of rewired interactions used for this calculation is also accurate and has taken into account for assay detection rates. Standard error was calculated using the delta method as described earlier. *P*-values were calculated using a *Z* test.

Our definition of rewiring is based on the interactions in the orthologous species. However, in cases where a paralog pair in the reference species shares an interaction that is rewired in the orthologous species, it is possible that the common ancestor may have this interaction. It could be argued that if the common ancestor does have the interaction, it is not truly neo-functionalized. To account for this (and since the interactome for the common ancestor is unknown), we constructed a set of interactions involving at least one protein that has a paralog, and the interactor of the paralog has only degree one (only one interaction), *i.e.*, by definition

that interactor cannot be shared between paralogs in the reference species. Even for this set, we find that *S. pombe* paralog pairs are significantly more neo-functionalized than *S. cerevisiae* paralog pairs, confirming the robustness of our results (Figure 3.10E).

3.4.31 Correcting for divergence times, sequence evolution rates and sequence identities

We obtained JTT-corrected divergence times for paralog pairs from Fares *et al*. (Fares et al., 2013). To ensure that the observed differences between SSD and WGD paralog pairs are not due to differences in divergence times, we selected only those SSD and WGD pairs whose divergence times are between the 10^{th} and the 90^{th} percentile of the WGD divergence time distribution. The rationale here is to use the WGD distribution as a reference (we remove the top and the bottom 10 percentiles to eliminate outliers) and sample SSD paralog pairs that are only from this divergence time window.

We used K_a to calculate sequence evolution rates. K_a (not K_s or K_a / K_s) is an appropriate choice to correct for sequence evolution rate because synonymous substitutions between WGD pairs are essentially saturated (Byrne and Wolfe, 2007). As mentioned earlier, to ensure that the observed differences between SSD and WGD paralog pairs are not due to differences in sequence evolution rates, we selected only those SSD and WGD pairs whose sequence evolution rate are between the 10^{th} and the 90^{th} percentile of the WGD K_a distribution.

We obtained paralog sequence identities from Ensembl BioMart. Here too, as earlier, to ensure that the observed differences between SSD and WGD paralog pairs are not due to differences in sequence identity, we selected only those SSD and WGD pairs whose identities are between the 10^{th} and the 90^{th} percentile of the WGD sequence identity distribution. Since sequence identity depends both on divergence time and sequence evolution rates, correcting for sequence identity simultaneously corrects for both covariates.

3.4.32 Functional properties of *S. cerevisiae* SSD and WGD pairs

To calculate the fraction of SSD and WGD pairs in complexes, we used high-quality literature curated complexes from CYC2008 (Pu et al., 2009). We computed the fractions of proteins from SGD and WGD pairs that are in all CYC2008 complexes, complexes with ≥10 proteins, and complexes with ≥20 proteins. *P*-values were calculated using a *Z* test.

To calculate the fraction of SSD and WGD pairs that involve non-essential genes but lead to synthetic lethality when both genes are deleted, we used genome-scale double knockout phenotype data (Costanzo et al., 2010). We considered a double deletion to lead to synthetic lethality if the genetic interaction score (ε) is strongly negative, *i.e.*, passes a stringent cutoff as defined by the authors at http://drygin.ccbr.utoronto.ca/~costanzo2009/ where ε<-0.12 and P<0.05. We considered a paralog pair to "share interactors" if both proteins had at least 2 interactors and they shared >50% of their interactors. "Other" paralog pairs are defined as those pairs that are not known to have any shared interactors based on the literature. *P*-values were calculated using a *Z* test.

To calculate the fraction of SSD and WGD pairs that are coexpressed, we used a normalized expression dataset constructed as described in Yu *et al*. (Yu et al., 2008). Paralog pairs that "share interactors" and "other" paralog pairs are defined as described above. We defined significant coexpression as *PCC* ≥a threshold value. To ensure that our conclusions are robust to the choice of this threshold, we used three different thresholds: 0.3, 0.4 and 0.5. When comparing the fractions of significantly co-expressed pairs, *P*-values were calculated using a *Z* test.

3.3.33 Calculation involving human SSD and WGD pairs

A set of human WGD (ohnolog) pairs was obtained from Makino and McLysaght (Makino and McLysaght, 2010). A set of human SSD pairs was identified as described in Singh *et al.* (Singh et al., 2014). This study also used the previous set of human WGD pairs (Makino and McLysaght, 2010) for their analyses. We calculated the fractions of SSD and WGD pairs containing genes that are known to cause the same disease based on HGMD (Stenson et al., 2014). Two genes are said to cause the same disease if at least one HGMD mutation on each of the two genes is associated with the same disease.

3.3.34 Direct Coupling Analysis for coevolutionary studies

To measure inter-protein evolutionary residue correlations, we performed coevolutionary analyses using DCA, which disentangles direct from indirect correlations among residue positions in evolutionarily-derived multiple sequence alignments (MSA). The most highly correlated residue pairs, as indicated by a high direct information (DI) score, can be used to predict contact residues, protein structures, and complex interfaces (Morcos et al., 2011). We compiled a list of orthologous protein sequences in *S. pombe*, *S. cerevisiae*, and 26 other yeast species by computing reciprocal best BLASTP hits between the proteome of *S. pombe* and the proteomes of these species accessed from UniProt (*82*). Cd-Hit was used to eliminate redundant sequences with >90% sequence identity from the list of orthologs for each *S. pombe* protein (Li and Godzik, 2006). MSAs among all remaining orthologs for each *S. pombe* protein were assembled using Clustal Omega (Sievers et al., 2011). For each studied *S. pombe* interaction known to be intact or coevolved in *S. cerevisiae*, we concatenated MSAs for the two proteins

involved and ran DCA using default parameters to find the inter-protein residue pairs with the highest correlations (DI scores). Homodimers were excluded from this calculation as it is impossible to disentangle intra from inter-protein evolutionary pressures. *P*-values were calculated using a *U* test.

3.5 RESULTS

3.5.1 A proteome-wide high-coverage binary protein interactome map of *S. pombe*

To generate a proteome-wide interactome network for *S. pombe*, which we call FissionNet, we systematically tested all pairwise combinations of proteins encoded by 4,989 *S. pombe* genes (corresponding to >99% of all *S. pombe* coding genes) using our high-quality yeast two-hybrid (Y2H) assay, the same pipeline that we used to generate the budding yeast and human interactome networks (Figure 3.2A) (Yu et al., 2008; Yu et al., 2011). Extensive screenings in triplicate (a total of ~75 million protein pairs) yielded 2,278 interactions between 1,305 proteins, of which 2,130 (93.5%) have not been previously reported in *S. pombe* (Figure 3.1A) (Das and Yu, 2012). Furthermore, FissionNet contains 1,034 interactions that have not been reported between orthologs in any other species before. Of these, 142 interactions involve *S. pombe* proteins that both have human orthologs, but at least one does not have a *S. cerevisiae* ortholog and, hence, cannot be studied in *S. cerevisiae*. Lastly, FissionNet consists of 917 (~40%) interactions in which an experimentally uncharacterized protein interacts with a protein of experimentally-verified known function. Thus, FissionNet provides a valuable repertoire of biological insights.

To assess the sensitivity and specificity of our Y2H assay (Yu et al., 2008), we constructed a positive reference set (PRS) consisting of 93 well-validated *S. pombe* interactions

from the literature and a negative reference set (NRS) of 168 random *S. pombe* protein pairs that are not known to interact in the literature and whose orthologs in other species are also not known to interact. We performed Y2H and protein complementation assay (PCA) (Das et al., 2013; Yu et al., 2008) to test what fraction of the PRS, NRS, and a random sample of 220 FissionNet interactions can be detected using orthogonal methods (Figure 3.1B). We found that the detection rates of the PRS and FissionNet interactions are indistinguishable from each other and are significantly higher than that of the NRS (Figure 3.1B; >15% difference in detection rates between the PRS and NRS for both assays, $P<10^{-3}$, *Z* test). The robust validation rates of FissionNet interactions by an orthogonal assay confirm the high quality of the network. Furthermore, although it has been speculated that Y2H interactions involving proteins with many interaction partners (hubs) could be of low quality (Bader et al., 2004), we found that the validation rate by PCA of hub interactions is the same as the overall PCA validation rate for FissionNet (Figure 3.1B; $P=0.34$, *Z* test), confirming that FissionNet interactions involving hubs are of high quality.

Biological relationships between interacting proteins in FissionNet were assessed by measuring similarities in protein localization, functional annotations, and expression profiles. We found that FissionNet interactions are significantly enriched for protein pairs that are co-localized, functionally similar, and encoded by coexpressed genes relative to random expectation (Figures 3.1C-E and 3.2B-H; $P<0.05$ in all three cases using a KS test for coexpression and *Z* test for co-localization and functional similarity). Furthermore, the enrichment of these interactions for all three categories is similar to that of literature-curated binary interactions. These results confirm that FissionNet interactions are functionally relevant *in vivo*. We illustrate

this by focusing on two previously unreported interactions: Tas3-Hhp1 and Atf1-Cid12, and their potential roles in gene silencing.

3.5.2 FissionNet provides insights into functions of proteins and interactions

The regulation of centromeric silencing is a well-conserved process in *S. pombe* and metazoans but is divergent from that in *S. cerevisiae* (Holoch and Moazed, 2015). FissionNet revealed a previously unidentified interaction between Tas3 and Hhp1 that we confirmed *in vivo* (Figures 3.1F-G). Tas3 is a component of the RNA-induced transcriptional silencing (RITS) complex that mediates gene silencing at *S. pombe* centromeres (Verdel et al., 2004). Hhp1 is a conserved mitotic checkpoint kinase (Johnson et al., 2013) not known to be involved in centromeric silencing. In *S. pombe* cells where the *ura4*$^{+}$ reporter gene was inserted at the centromere inner repeats of chromosome 1 (*imr1R*) (Verdel et al., 2004), we find that *hhp1Δ* confers loss of silencing at the centromere, similar to *tas3Δ* cells (Figure 3.1H). Furthermore, levels of endogenous centromeric transcripts are elevated in *hhp1Δ* cells (Figure 3.2I). Moreover, loss of *hhp1* leads to a decrease in the dimethylation of histone 3 lysine 9 (H3K9me2) at the centromere (Figure 3.2J). These results show that Hhp1 is involved in centromeric silencing.

We also identified a previously unreported interaction between the transcription factor Atf1 and the polyadenylation polymerase Cid12 (Figure 3.3A). Atf1 mediates transcriptional responses to stresses such as high temperatures (Shiozaki and Russell, 1996). At *S. pombe* centromeres, Cid12 is a core component of the RNA-directed RNA-polymerase complex (RDRC) (Motamedi et al., 2004). The RDRC is responsible for generating double-stranded RNAs, a key step for Dcr1-dependent centromeric silencing. Interestingly, it has been reported

that Dcr1 transcriptionally represses the Atf1-target genes *hsp16* and *hsp104* under non-stressed conditions (Woolcock et al., 2012).

Pull-down experiments confirm the interaction of Atf1 and Cid12 in *S. pombe* (Figure 3.3B), and *cid12Δ* cells grown under non-stressed conditions show elevated mRNA levels of *hsp16* and *hsp104* as compared to wild-type cells, similar to *dcr1Δ* cells (Figure 3.3C). Additionally, double mutant *cid12Δ dcr1Δ* cells do not exhibit more drastic transcript accumulation than the single deletion mutants, suggesting both genes function in the same pathway (Figure 3.3C). Together, these results suggest that Cid12 may be involved in repressing aberrant gene expression of Atf1-target genes.

Next, we identified two Cid12 mutations, lysine-213 to isoleucine ($Cid12^{K213I}$) and aspartic acid-260 to valine ($Cid12^{D260V}$), that disrupt the interaction of Cid12 with Atf1 while preserving interactions within the RDRC complex (Figure 3.3D). Exogenous expression of wild-type Cid12 in *cid12Δ* cells enables the transcriptional repression of *hsp16* and *hsp104*. In stark contrast, neither mutant can repress gene expression (Figure 3.3E). The mutant phenotype is not due to complete loss of protein caused by destabilization because these Cid12 mutant proteins express in *S. pombe* cells (Figure 3.4A). Furthermore, in *cid12Δ* cells where the $ura4^+$ reporter gene was inserted at the centromeric *imr1R*, we find that exogenous expression of either Cid12 wild-type or mutants equally permit the silencing of the $ura4^+$ reporter (Figure 3.3F). Thus, we show that Cid12 has dual roles in regulating the expression of heat-shock genes and the centromere. Importantly, the roles can be selectively uncoupled via specific disruption of the Atf1-Cid12 interaction. These examples illustrate the usefulness of FissionNet as a resource to uncover areas of biological inquiry.

3.4.3 Comparative network analyses reveal species-specific conservation of interactions

High-quality protein interactome networks have previously been reported in budding yeast (Yu et al., 2008) and human (Rolland et al., 2014). A fundamental question, which can be addressed with FissionNet and these networks, is how protein-protein interactions have evolved and whether this trend mirrors gene-level evolution. From sequence-based phylogenetic analyses, the two yeasts are less divergent from each other than either yeast is from human (Figure 3.5A) (Sipiczki, 2000). Additionally, the two yeasts share a greater fraction of protein-coding genes than either yeast does with human (Figures 3.6A-B).

To calculate interaction conservation, we considered only those interactions that have the potential to be conserved, *i.e.*, the two interacting proteins in the reference species have orthologs in the other species. However, directly calculating the overlap between sets of interactions obtained from the literature would be erroneous because currently available interactomes are incomplete and are derived from assays with varied and often unreported false positive and false negative rates (Yu et al., 2008). Therefore, to accurately estimate the underlying interaction conservation fractions, we required interactomes of all species to be derived from the same experimental assay. Since interactomes in budding yeast (Yu et al., 2008) and human (Rolland et al., 2014) have been generated using our version of Y2H (Figure 3.6C-D), we were able to compare FissionNet to these interactome networks to measure the observed extent of interaction conservation. We developed a rigorous Bayesian framework that incorporates both the false positive and false negative rates of our Y2H assay to estimate the underlying interaction conservation fraction from the observed fraction for each pair of species. Surprisingly, we find that interaction conservation follows a completely different trend from gene conservation (Figures 3.5B and 3.6E-F). While only ~40% of *S. pombe* interactions are conserved in *S.*

Figure 3.1 A Proteome-wide Binary Protein Interactome Map of *S. pombe*

(A) Network representation of FissionNet. Proteins are color-grouped based on PomBase GO slim categories. The number of FissionNet interactions per group is indicated. (B) Y2H and PCA detection rates of the PRS, NRS, FissionNet, and FissionNet hub interactions. (C) Pearson correlation coefficient (*PCC*) distribution of gene expression profiles of interacting and all random protein pairs. (D) Enrichment of co-localized protein pairs. (E) Enrichment of protein pairs sharing similar functions. (F) Subnetwork of Tas3 and Hhp1 in FissionNet. (G) Coimmunoprecipitation of Tas3-myc and Hhp1-HA *in vivo*. (H) Centromeric silencing assay of *tas3Δ* and *hhp1Δ* cells. A schematic of the *imr1R* region with the *ura4*$^{+}$ reporter gene is shown. WT denotes wild-type. Data are shown as measurements + standard error (SE). * denotes significant (P<0.05); n.s. denotes not significant.

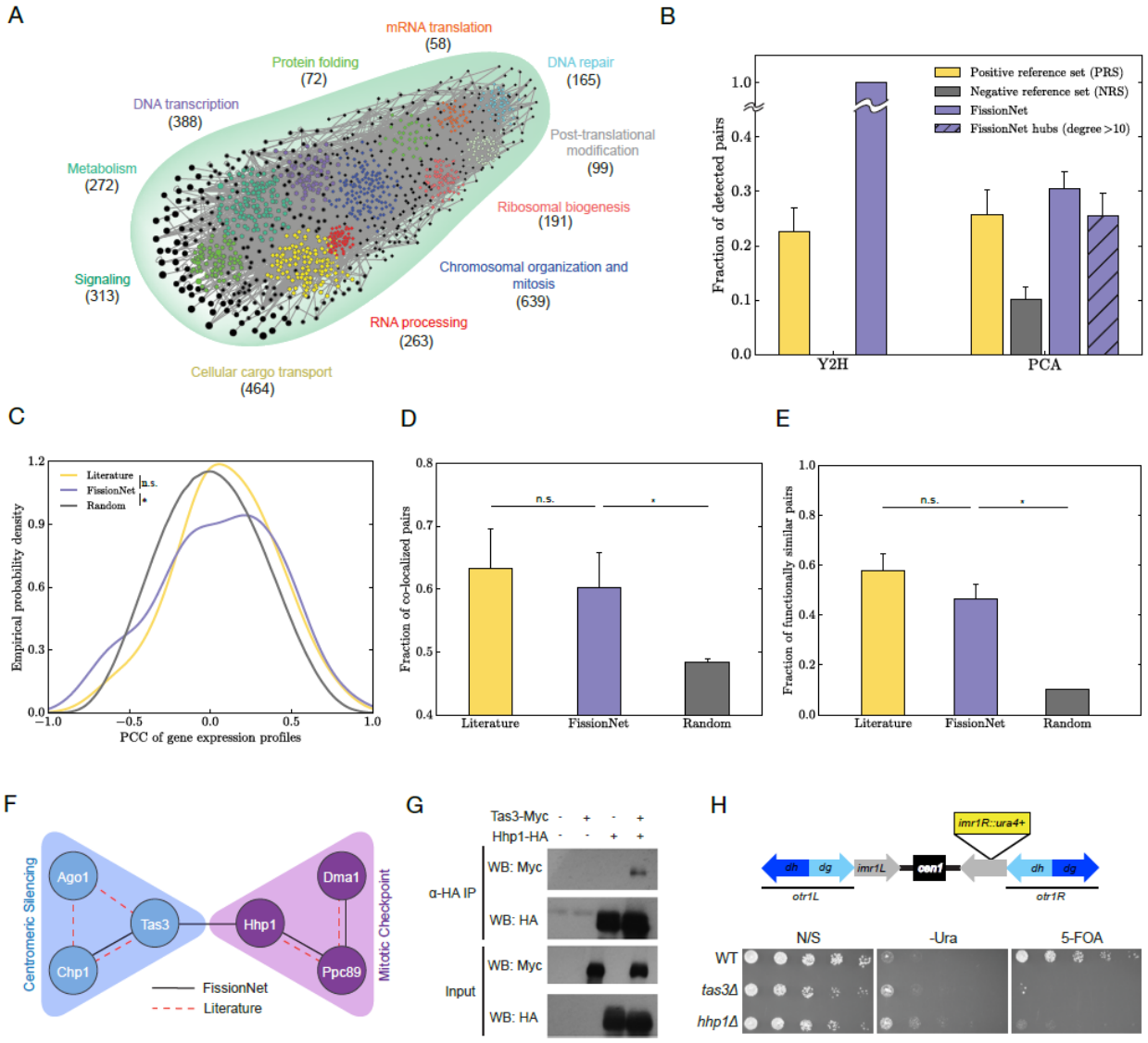
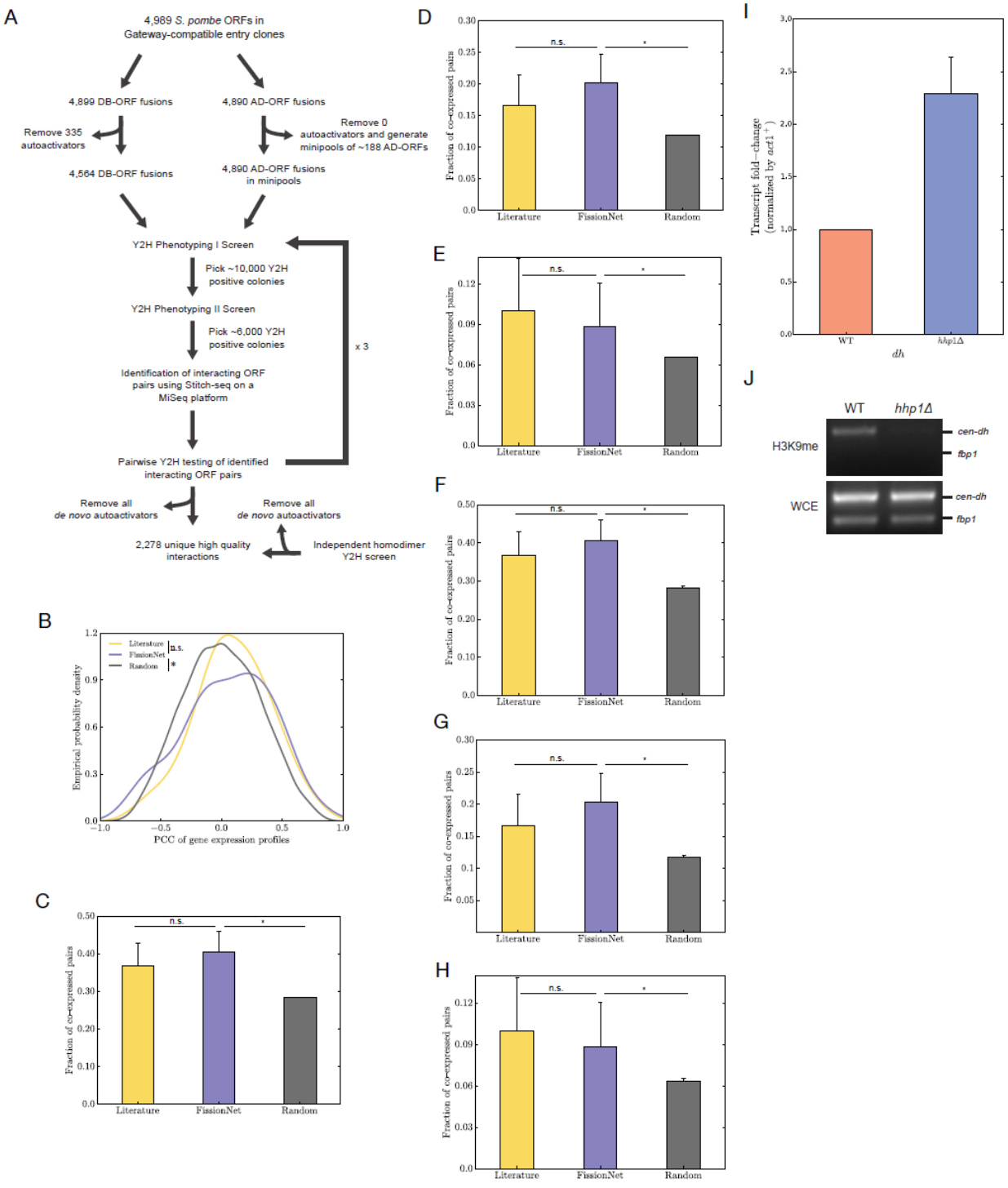

A
mRNA translation
(58)
Protein folding
(72)
DNA repair
(165)
DNA transcription
(388)
Post-translational
modification
(99)
Metabolism
(272)
Ribosomal biogenesis
(191)
Signaling
(313)
Chromosomal organization and
mitosis
(639)
RNA processing
(263)
Cellular cargo transport
(464)
B
Positive reference set (PRS)
Negative reference set (NRS)
FissionNet
FissionNet hubs (degree >10)
Fraction of detected pairs
Y2H
PCA
C
Literature
FissionNet
Random
n.s.
*
Empirical probability density
PCC of gene expression profiles
D
Fraction of co-localized pairs
Literature
FissionNet
Random
E
Fraction of functionally similar pairs
Literature
FissionNet
Random
F
Centromeric Silencing
Mitotic Checkpoint
Ago1
Tas3
Chp1
Hhp1
Dma1
Ppc89
FissionNet
Literature
G
Tas3-Myc
Hhp1-HA
α-HA IP
WB: Myc
WB: HA
Input
WB: Myc
WB: HA
H
imr1R::ura4+
dh
dg
imr1L
cen1
otr1L
otr1R
N/S
-Ura
5-FOA
WT
tas3Δ
hhp1Δ

Figure 3.2 FissionNet is a High Quality Interactome Network

(A) Schematic representation of our Y2H pipeline to generate FissionNet. (B) Pearson correlation coefficient (*PCC*) distribution of gene expression profiles of interacting and random protein pairs, where random pairs are generated by permuting edges in FissionNet. (C-E) Fraction of significantly coexpressed protein pairs from the literature, FissionNet, and among all random pairs where significant coexpression is defined as (C) *PCC*>0.2, (D) *PCC*>0.4 and (E) *PCC*>0.5 respectively. (F-H) Fraction of significantly coexpressed protein pairs from the literature, FissionNet, and among random pairs generated by permuting edges in FissionNet where significant coexpression is defined as (F) *PCC*>0.2, (G) *PCC*>0.4 and (H) *PCC*>0.5. (I) Quantitative real-time PCR (qRT-PCR) confirms elevated levels of centromeric *dh* in *hhp1Δ* cells. (J) ChIP shows loss of H3K9me2 at centromeric *dh* chromatin in *hhp1Δ* cells. WT denotes wild-type. Data are shown as measurements + standard error (SE). * denotes significant (P<0.05); n.s. denotes not significant. $Act1^{+}$ serves as loading control for qRT-PCR experiment. *Fbp1* serves as a euchromatin control locus.


A
4,989 S. pombe ORFs in Gateway-compatible entry clones
4,899 DB-ORF fusions
4,890 AD-ORF fusions
Remove 335 autoactivators
Remove 0 autoactivators and generate minipools of ~188 AD-ORFs
4,564 DB-ORF fusions
4,890 AD-ORF fusions in minipools
Y2H Phenotyping I Screen
Pick ~10,000 Y2H positive colonies
Y2H Phenotyping II Screen
Pick ~6,000 Y2H positive colonies
x 3
Identification of interacting ORF pairs using Stitch-seq on a MiSeq platform
Pairwise Y2H testing of identified interacting ORF pairs
Remove all de novo autoactivators
Remove all de novo autoactivators
2,278 unique high quality interactions
Independent homodimer Y2H screen
B
Literature
FissionNet
Random
n.s.
*
Empirical probability density
PCC of gene expression profiles
C
Fraction of co-expressed pairs
Literature
FissionNet
Random
D
E
F
G
H
I
Transcript fold-change (normalized by act1+)
WT
hhp1Δ
dh
J
WT
hhp1Δ
H3K9me
WCE
cen-dh
fbp1

cerevisiae (of the 1,331 interactions where both proteins have *S. cerevisiae* orthologs and were pairwise retested using our Y2H assay), ~65% of *S. pombe* interactions are conserved in human (of the 652 interactions where both proteins have human orthologs and were pairwise retested using our Y2H assay) (Figure 3.5B; $P=1.4\times10^{-4}$, *Z* test). However, when using budding yeast as the reference species, the fraction of conserved interactions is as high in human as in *S. pombe*, comparable to the fraction conserved between *S. pombe* and human (Figure 3.5B). Moreover, when using human as the reference species, the fraction of conserved interactions is higher in *S. pombe* (~65%) than in *S. cerevisiae* (~42%). We were able to recapitulate these results using interaction datasets generated by other assays (Figures 3.6G-I; >1.5 fold difference between fission yeast interactions conserved in budding yeast and human; $P<10^{-3}$ in all cases, *Z* test). Thus, our results suggest that a large fraction of interactions are conserved between human and *S. pombe*, but have been lost specifically in the *S. cerevisi*ae lineage.

One possible explanation for these surprising results is that fission yeast proteins that are conserved in human could have higher overall sequence similarity than those that are conserved in budding yeast. However, we find that proteins in interactions that have the potential to be conserved based on orthology are actually slightly more similar in sequence between the two yeasts than between *S. pombe* and human (Figure 3.5C; $P<10^{-5}$, *U* test).

Another possibility is that the observed difference primarily arises from interactions involving proteins that are conserved between fission yeast and human but lost in budding yeast. To test this, we first focused on proteins that are conserved in all 3 species. We still find that ~20% more interactions are conserved between *S. pombe* and human as compared to between the two yeasts (Figures 3.5D and 3.6J-K; $P<0.05$, *Z* test).

We next explored the conservation of interactions involved in various biological

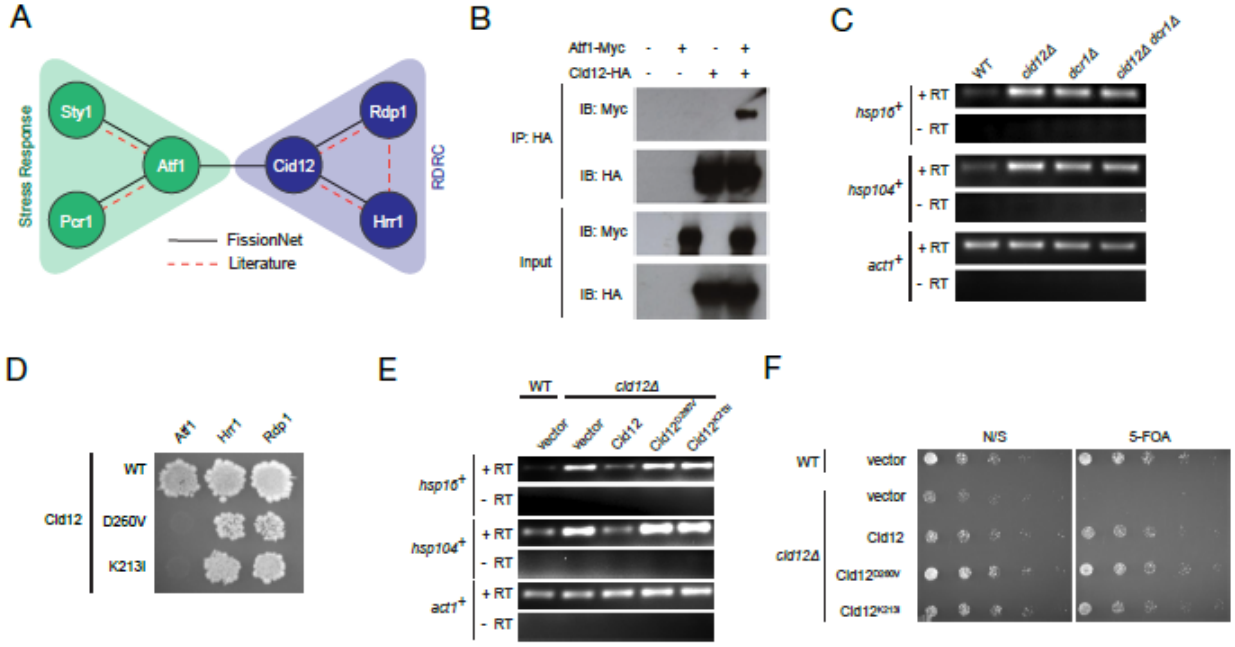


Figure 3.3 Atf1-Cid12 Interaction Mediates Silencing at Heat-shock Genes

(A) Subnetwork of Atf1 and Cid12 in FissionNet. (B) Coimmunoprecipitation of Atf1-myc and Cid12-HA *in vivo*. (C) Semi-quantitative real-time PCR (semi qRT-PCR) shows *hsp16* and *hsp104* transcript levels in deletion strains. (D) Y2H confirms Cid12 mutants cannot interact with Atf1, but maintain interactions with Hrr1 and Rdp1. (E) Semi qRT-PCR shows that the Cid12 mutants in *cid12Δ* cells do not restore the repression of *hsp16* or *hsp104*. (F) Centromeric silencing assay shows that Cid12 mutants retain centromeric silencing function. -RT, no reverse transcriptase. +RT, with reverse transcriptase. $Act1^{+}$ serves as loading control. WT denotes wild-type.

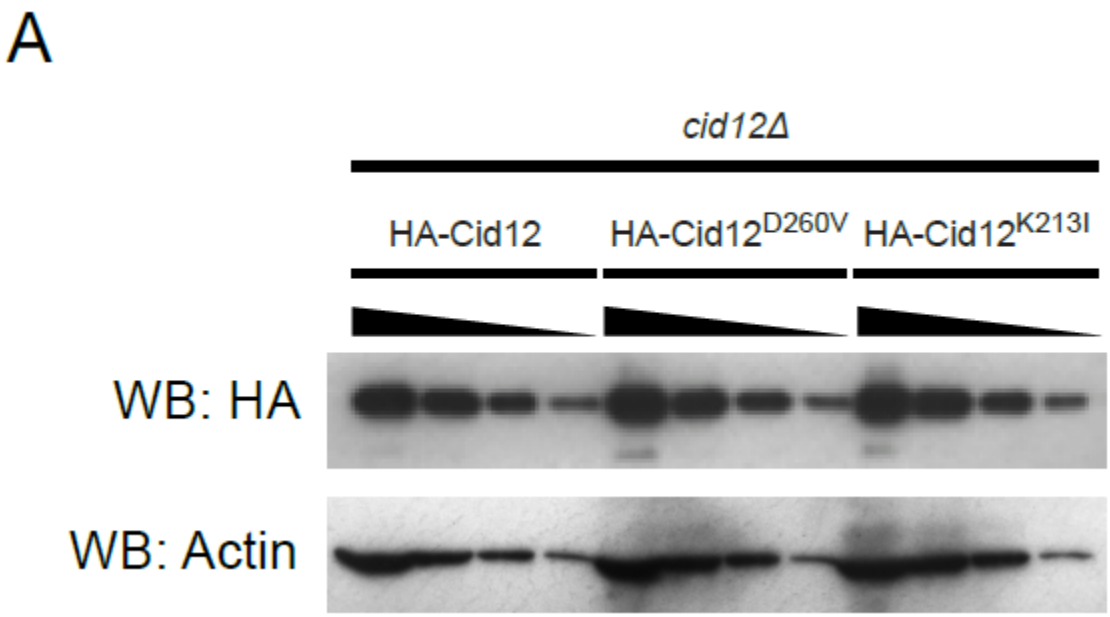


Figure 3.4 Mutant Cid12 Protein Expression Levels in *S. pombe* Cells

(A) Western blot shows protein expression levels of Cid12-HA wild-type or mutants. Blot was performed using titrated amounts of total protein. Four-fold serial dilution was performed for each protein. Actin serves as loading control. Blot performed by Xiaomu Wei.

processes as defined by the Gene Ontology (GO) (Ashburner et al., 2000). We find wide variation in species-specific interaction conservation among different processes (Figures 3.5E and 3.6L-N). We show that *S. pombe* interactions are more conserved in human than in *S. cerevisiae* for 10 out of 13 GO Slim categories containing ≥50 interactions (Figure 3.5E; $P<0.05$, as marked, *Z* test). The same trend is observed with GO Slim categories containing ≥30 or ≥75 interactions (Figures 3.6L and 3.6N). Some of these categories, such as "chromosomal organization", "chromosome segregation", and "cell cycle", are far better conserved in human than in *S. cerevisiae*, and accordingly *S. pombe* has been used as a model organism for studying these processes (Wood et al., 2002). Furthermore, considering GO Slim categories that are well conserved in all three species (using cutoffs of ≥50, 100, and 200 genes annotated per species), we find that the conservation of *S. pombe* interactions in these core biological processes is also higher in human than in *S. cerevisiae* (Figures 3.5F and 3.6O; $P<10^{-3}$, *Z* test). Overall, these results suggest that insights gained from FissionNet may be widely applicable to the study of human biology across many important cellular processes.

We validated three cases of previously unreported functional conservation between fission yeast and human proteins. Uncharacterized *S. pombe* factors Srrm1, SPAC30D11.14C, and SPAC1952.06C interact with known splice factors Srp1, Usp104, and Cwf15, respectively (Figure 3.5G**)**. Although these proteins have no orthologs in *S. cerevisiae*, they are orthologous to human SRRM1, KIAA0907, and CTNNBL1, respectively. Interestingly, all three human orthologs have been implicated in pre-mRNA splicing or were found to associate with spliceosomal factors in human (Blencowe et al., 1998; Hegele et al., 2012; Rolland et al., 2014). We used DNA microarrays to measure changes in the splicing of every known intron in the *S. pombe* deletion mutants. The loss of *srrm1*, *SPAC30D11.14C*, or *SPAC1952.06C* results in

widespread splicing defects, confirming the roles for these proteins in the splicing pathway (Figure 3.5H). Moreover, Srrm1 and Srp1 share many gene targets, suggesting that the interacting proteins are functionally related (Figure 3.6P). Notably, an analysis of the introns whose splicing is affected by *srrm1* deletion shows a strong enrichment for introns with weak splice site signals (Figure 3.6Q). This is consistent with previous findings that human *SRRM1* affects splice site selection by binding to exonic splicing enhancers and facilitating interactions between spliceosomal proteins (Blencowe et al., 1998). These results highlight the utility of FissionNet to reveal proteins that are functionally conserved between *S. pombe* and human.

3.5.4 Determinants of interaction conservation

Previous studies have shown that increased protein sequence similarity facilitates conservation of protein interactions (Matthews et al., 2001). Indeed, we also found a positive correlation between sequence similarity of proteins and the fraction of their associated interactions conserved between *S. pombe* and human or *S. cerevisiae*, demonstrating a proteome-scale dependence of protein sequence and function (Figure 3.7A; $R^2_{S.p\text{-}H.s}$=0.948 and $R^2_{S.p\text{-}S.c}$=0.976). However, protein interaction conservation is not completely dependent on overall sequence similarity, as we find many instances of conserved interactions involving proteins with low overall sequence similarity (<40%) with their orthologs (Figure 3.7A; 40% and 13% of 116 interactions in human and 196 interactions in *S. cerevisiae*, respectively). To investigate whether certain highly conserved domains in these proteins play an important role in interaction conservation, we inferred protein interaction domains from co-crystal structures of 124 human interactions conserved in *S. pombe* and 293 conserved in *S. cerevisiae*. We find that the sequence similarity within protein interaction domains tends to be higher than in other domains for

interactions conserved between fission yeast and human (Figure 3.7B; 7.0% higher, P=0.012, U test). For instance, the human DR1-DRAP1 heterodimer is orthologous to the protein pair Ncb2 and Dpb3 in *S. pombe*. While the overall sequence similarity of the orthologs is quite low (0.58 and 0.51, respectively), the interaction is conserved in fission yeast. Moreover, we also find that the proteins can interact with the orthologs of their native interaction partner (Figure 3.7C). Based on a crystal structure of the human DR1-DRAP1 complex, we were able to determine the interaction domains of these proteins (Figure 3.7D) (Kamada et al., 2001). The sequence similarity within these domains in DR1 and DRAP1 with their fission yeast orthologs is 0.78 and 0.80, respectively, while the conservation outside of these interaction domains is only 0.45 and 0.38. Thus, the basis for this high degree of functional conservation is likely dependent on the interaction domains.

Strikingly, interaction conservation is nearly three times higher between *S. pombe* and human than between the two yeasts at low levels of overall sequence similarity (Figure 3.7A; at <40% similarity, P=0.030, Z test). As sequence similarity approaches 100%, interaction conservation converges. Therefore, for the vast majority of interactions corresponding to proteins with lower sequence similarity to their orthologs, our results strongly suggest that species-specific factors, independent of overall protein sequence similarity, influence conservation of protein-protein interactions.

We then sought to explore other factors that could explain the basis of interaction conservation. First, we used ClusterOne (Nepusz et al., 2012) to detect topological protein clusters in FissionNet. We find that intra-cluster FissionNet interactions are >3 times more likely to be conserved in both budding yeast and human than inter-cluster interactions (Figures 3.7E and 3.8A; P<0.05 for both organisms, Z test). Next, we examined biological processes defined

Figure 3.5 *S. pombe* Protein Interactions are More Conserved in Human than in *S. cerevisiae* (A) Sequence-based phylogeny dendrogram of *S. pombe* (*S.p.*), *S. cerevisiae* (*S.c.*), and human (*H.s.*). (B) Interaction conservation between reference-query species. (C) Sequence conservation for ortholog pairs that could be conserved between *S.p.-S.c.* and *S.p.-H.s.* (D) Interaction conservation between reference-query species for proteins that are conserved in all three species. (E) Interaction conservation in GO Slim categories with at least 50 interactions. (F) Interaction conservation among GO Slim categories that are conserved in all three species. (G) FissionNet subnetworks of Srrm1, SPAC30D11.14C, and SPAC1952.06C. (H) Global splicing profiles of deletion strains relative to wild-type. Columns represent total mRNA (T), pre-mRNA (P), and mature mRNA (M). Data are shown as measurements + SE. * denotes significant (P<0.05); n.s. denotes not significant. Panels A-F made by Jishnu Das and Michael Meyer.

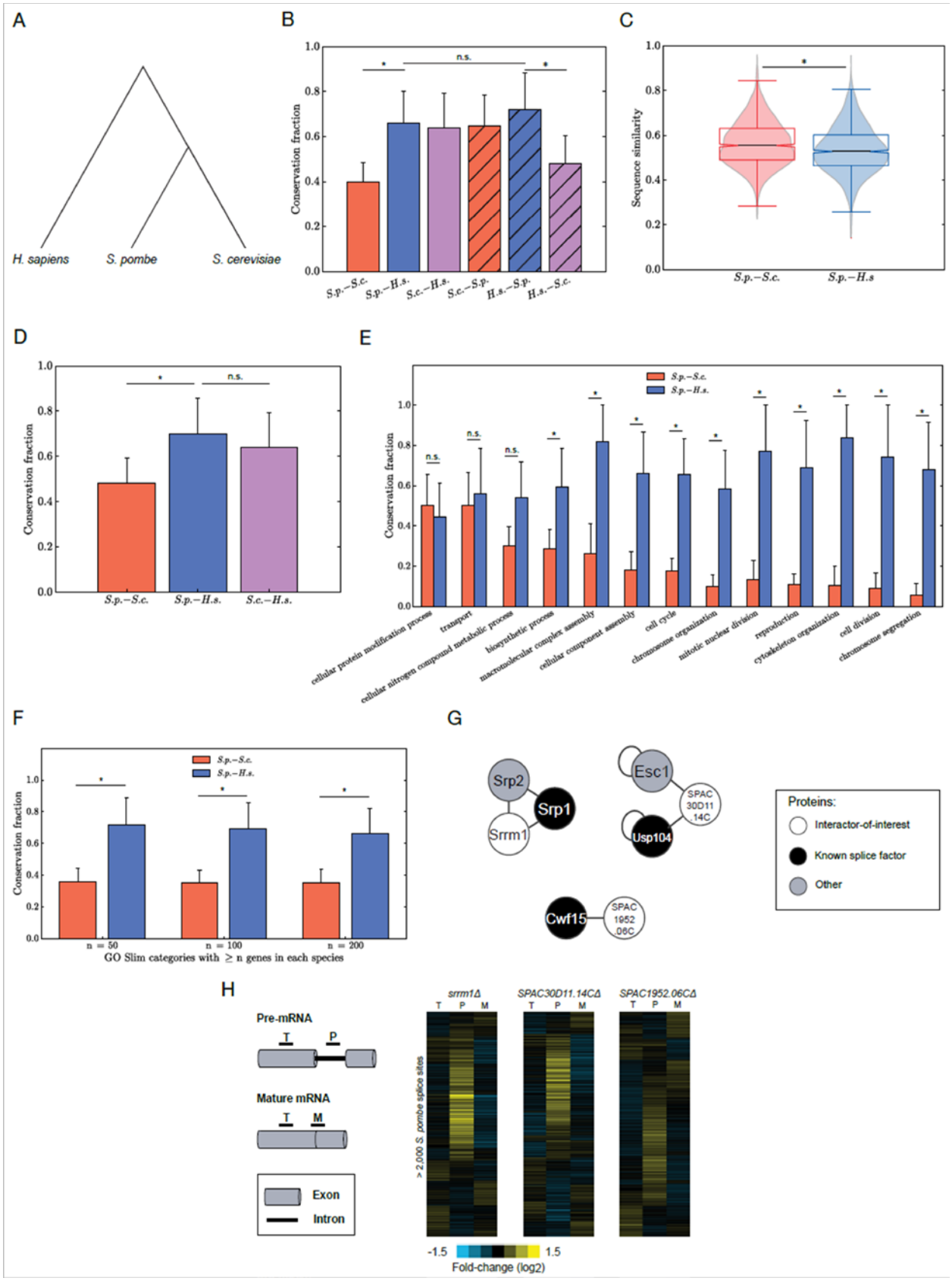
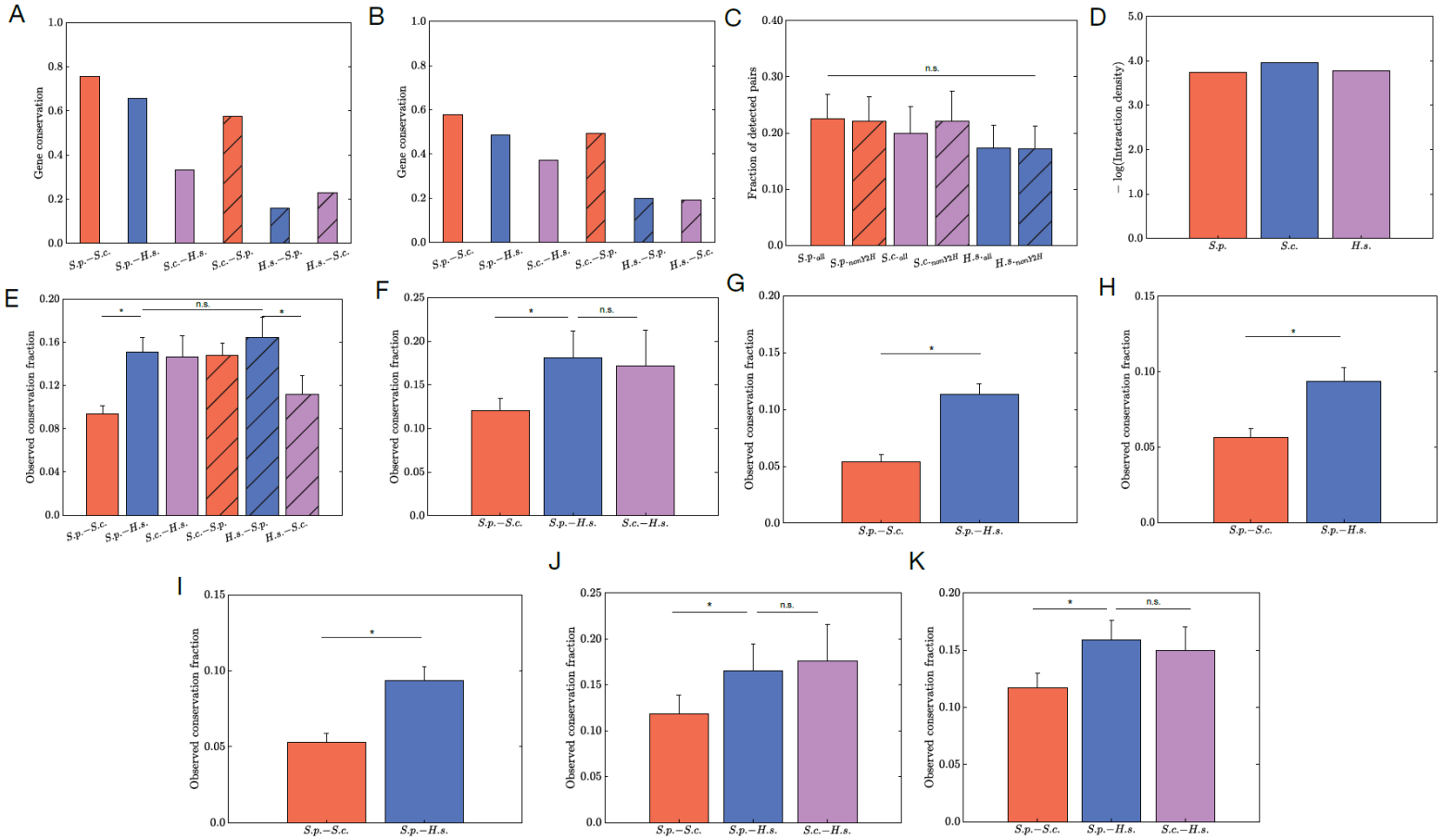

A
H. sapiens
S. pombe
S. cerevisiae
B
Conservation fraction
n.s.
S.p.–S.c.
S.p.–H.s.
S.c.–H.s.
S.c.–S.p.
H.s.–S.p.
H.s.–S.c.
C
Sequence similarity
S.p.–S.c.
S.p.–H.s
D
Conservation fraction
S.p.–S.c.
S.p.–H.s.
S.c.–H.s.
E
Conservation fraction
S.p.–S.c.
S.p.–H.s.
cellular protein modification process
transport
cellular nitrogen compound metabolic process
biosynthetic process
macromolecular complex assembly
cellular component assembly
cell cycle
chromosome organization
mitotic nuclear division
reproduction
cytoskeleton organization
cell division
chromosome segregation
F
Conservation fraction
S.p.–S.c.
S.p.–H.s.
n = 50
n = 100
n = 200
GO Slim categories with ≥ n genes in each species
G
Srp2
Srp1
Srrm1
Esc1
SPAC 30D11 .14C
Usp104
Cwf15
SPAC 1952 .06C
Proteins:
Interactor-of-interest
Known splice factor
Other
H
Pre-mRNA
T
P
Mature mRNA
T
M
Exon
Intron
srrm1Δ
SPAC30D11.14CΔ
SPAC1952.06CΔ
T P M
> 2,000 S. pombe splice sites
-1.5
1.5
Fold-change (log2)

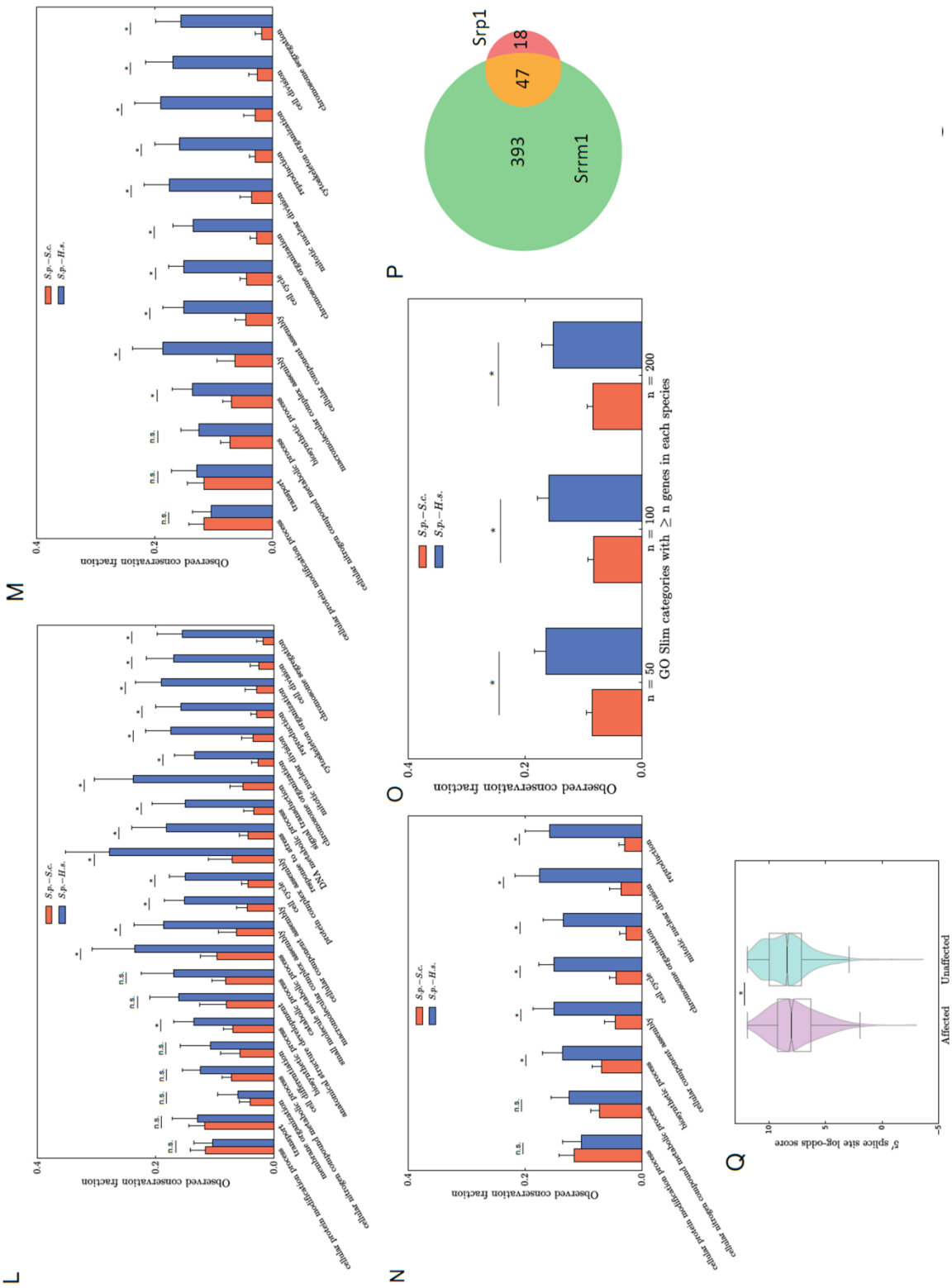
Figure 3.6 *S. pombe* Protein Interactions are More Conserved in Human than *S. cerevisiae*

(A-B) Pairwise comparisons of the conservation of all coding genes between reference-query species pairs. (A) Orthologs are defined by PomBase (McDowall et al., 2015) and Saccharomyces Genome Database (SGD) (Cherry et al., 2012). (B) Orthologs are defined by InParanoid (Sonnhammer and Ostlund, 2015). (C) Y2H detection rates of PRS and PRS_nonY2H (subset of PRS interactions that have been detected using an assay other than Y2H) interactions in fission yeast, budding yeast and human. (D) Interaction density, *i.e.*, interactions detected out of the total number of proteins pairs screened (log scale) in different organisms. (E) Observed interaction conservation between reference-query species pairs. (F) Observed interaction conservation between reference-query species pairs for proteins that have 1:1 orthologs between reference and query species. (G) Observed interaction conservation between reference-query species pairs using co-crystal structures for *S. cerevisiae* and human. (H-I) Observed interaction conservation between reference-query species pairs using large-scale AP/MS datasets for *S. cerevisiae* and human. For both panels, the human AP/MS dataset used is from (Huttlin et al., 2015). (H) The *S. cerevisiae* AP/MS dataset is from (Gavin et al., 2006). (I) The *S. cerevisiae* AP/MS dataset is from (Krogan et al., 2006). (J) Observed interaction conservation between reference-query species pairs for proteins that have 1:1:1 orthologs between fission yeast, budding yeast, and human. (K) Observed interaction conservation between reference-query species pairs for proteins that are conserved in all eukaryotes. (L-N) Observed conservation fractions of *S. pombe* interactions in *S. cerevisiae* and human in different GO Slim biological process categories with at least (L) 30, (M) 50, and (N) 75 interactions. (O) Observed interaction conservation among GO Slim categories that are conserved in all three species. (P) Overlap of genes whose intron splicing is affected by deletion of either Srrm1 or its interaction

partner Srp1. Indicated within the diagrams are the number of genes affected. (Q) Distribution of log odds scores of affected (intron accumulation of $\log_2$ 0.5 or greater in *srrm1Δ* versus wild type cells) versus unaffected (intron accumulation of less than $\log_2$ 0.5 in *srrm1Δ* versus wild type cells). The log odds score for each annotated 5' splice site measures the sequence similarity of that site relative to the consensus 5' splice site of each intron. Data are shown as measurements + SE. * denotes significant (P<0.05); n.s. denotes not significant. Analyses performed by Jishnu Das and Michael Meyer.


A
Gene conservation
S.p.−S.c.
S.p.−H.s.
S.c.−H.s.
S.c.−S.p.
H.s.−S.p.
H.s.−S.c.
B
Gene conservation
C
Fraction of detected pairs
n.s.
D
− log(Interaction density)
S.p.
S.c.
H.s.
E
Observed conservation fraction
F
Observed conservation fraction
G
Observed conservation fraction
H
Observed conservation fraction
I
Observed conservation fraction
J
Observed conservation fraction
K
Observed conservation fraction

(Figure 3.6 continued)



by GO (Ashburner et al., 2000) and observed the same trend (Figures 3.7F and 3.8B; $P<10^{-3}$ for both organisms, *Z* test). Using genetic interactions, it has been earlier hypothesized that while individual functional modules are conserved, inter-modular connectivity could be rewired across evolution (Roguev et al., 2008). In this study, we provide direct molecular level evidence on a proteome scale that while interactions within modules tend to be conserved across evolution, the cross-talk among these modules changes significantly from one species to another.

3.5.5 Gene duplication shapes the functional fate of paralogs

Gene duplication has long been known as a major source of evolutionary novelty (Arabidopsis Interactome Mapping Consortium, 2011). Previous studies have found that a whole-genome duplication (WGD) event leads to more functional redundancy between paralogous proteins than small-scale duplications (SSDs) (Arabidopsis Interactome Mapping Consortium, 2011; Hakes et al., 2007b). However, there has been much debate in the literature regarding the relative extents of sub-functionalization and neo-functionalization for diverged paralogs (Gibson and Goldberg, 2009; He and Zhang, 2005). Previous studies on functional evolution of paralogs often used interaction datasets from the literature, which, as mentioned earlier, suffer from detection and completeness biases (Yu et al., 2008). Until now, it has not been possible to measure the extent of sub-functionalization and neo-functionalization using an unbiased framework because there was only one proteome-wide high-coverage binary protein interactome available, that of *S. cerevisiae*. Here, we compare the unbiased proteome-wide networks of *S. pombe* (FissionNet) and *S. cerevisiae* (CCSB-YI1) (Yu et al., 2008) that we produced using the same Y2H assay to analyze these two types of functional divergence.

We first examined the extent to which interactions in *S. pombe* and *S. cerevisiae* tend to be conserved across species but not shared between within-species paralogs (sub-functionalized)

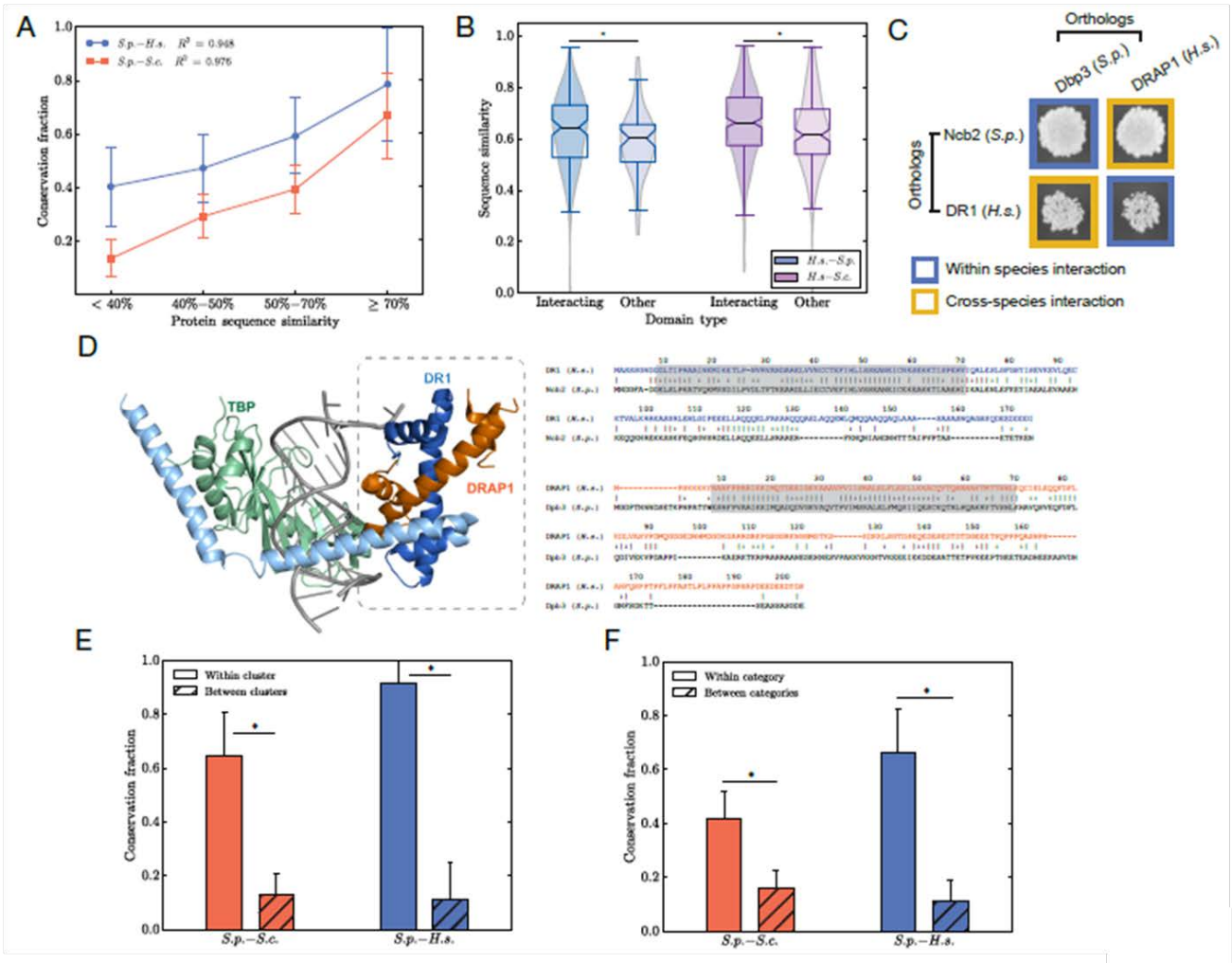
(Figure 3.9A). We find that fission yeast paralog pairs tend to undergo more sub-functionalization than budding yeast paralog pairs (Figure 3.9B; difference in log odds ratio=2.8 using 1,762 fission yeast paralog pairs and 2,068 budding yeast paralog pairs, $P<10^{-5}$, *Z* test). Since *S. pombe* paralogs arose via SSDs, while many *S. cerevisiae* paralogs arose via a WGD event (Kellis et al., 2004), this result suggests that duplication modes could impact paralog divergence differently. To test this, we compared paralog pairs generated via the WGD event with those generated via SSDs in *S. cerevisiae*. We find that SSD pairs are more sub-functionalized than WGD pairs (Figures 3.9C and 3.10A-D; $P<0.05$, *Z* test).

Next, we compared the extent of neo-functionalization (rewiring) (Figure 3.9A) for the two species and found that fission yeast paralog pairs tend to undergo more neo-functionalization than budding yeast pairs (Figures 3.9D and 3.10E; difference in log odds ratio=0.5 using 1,158 fission yeast paralog pairs and 1,175 budding yeast paralog pairs, $P=0.015$, *Z* test). Furthermore, within *S. cerevisiae*, SSD pairs are significantly more neo-functionalized than WGD pairs (Figures 3.9E and 3.10F-I; $P<0.05$, *Z* test).

In a WGD, the entire genome is duplicated almost at once. Soon afterward, a vast majority of the duplicates are purged while only a few are retained (Kellis et al., 2004). However, the duplicates that remain are under strong evolutionary pressure to maintain stoichiometric ratios with their interaction partners and, thus, evolve more slowly (Fares et al., 2013). On the other hand, SSDs arise sporadically and are under less pressure to maintain stoichiometric ratios (Fares et al., 2013), which explains why they can undergo greater functional divergence. This increased pressure on WGD genes to maintain stoichiometry is illustrated by their propensity to be enriched in protein complexes compared to SSD genes (Hakes et al., 2007b). Using 408 high-quality literature-curated complexes from CYC2008 (Pu et al., 2009),

Figure 3.7 Determinants of Interaction Conservation

(A) Interaction conservation as a function of overall protein sequence similarity. (B) Sequence similarity within protein interaction domains and other domains for interactions conserved between yeasts and human. (C) Y2H confirms the interactions of human (*H.s.*) DRAP1-DR1, the orthologous *S. pombe* (*S.p.*) Dpb3-Ncb2, and the cross-species interactions. (D) Crystal structure of human DR1-DRAP1. Boxed region highlights interaction domains. Gray shaded regions denote aligned interaction domain sequences. (E) Interaction conservation within and across topological clusters. (F) Interaction conservation within and across GO categories. Data are shown as measurements + SE. * denotes significant ($P<0.05$); n.s. denotes not significant. Abbreviations are *S. pombe* (*S.p.*), *S. cerevisiae* (*S.c.*), and human (*H.s.*). Panels A, B, D, E, and F made by Jishnu Das and Michael Meyer.


A
B
C
D
E
F
Conservation fraction
Protein sequence similarity
< 40%
40%–50%
50%–70%
≥ 70%
Sequence similarity
Interacting
Other
Domain type
Orthologs
Dbp3 (S.p.)
DRAP1 (H.s.)
Ncb2 (S.p.)
DR1 (H.s.)
Within species interaction
Cross-species interaction
TBP
DR1
DRAP1
Within cluster
Between clusters
Within category
Between categories
S.p.–S.c.
S.p.–H.s.

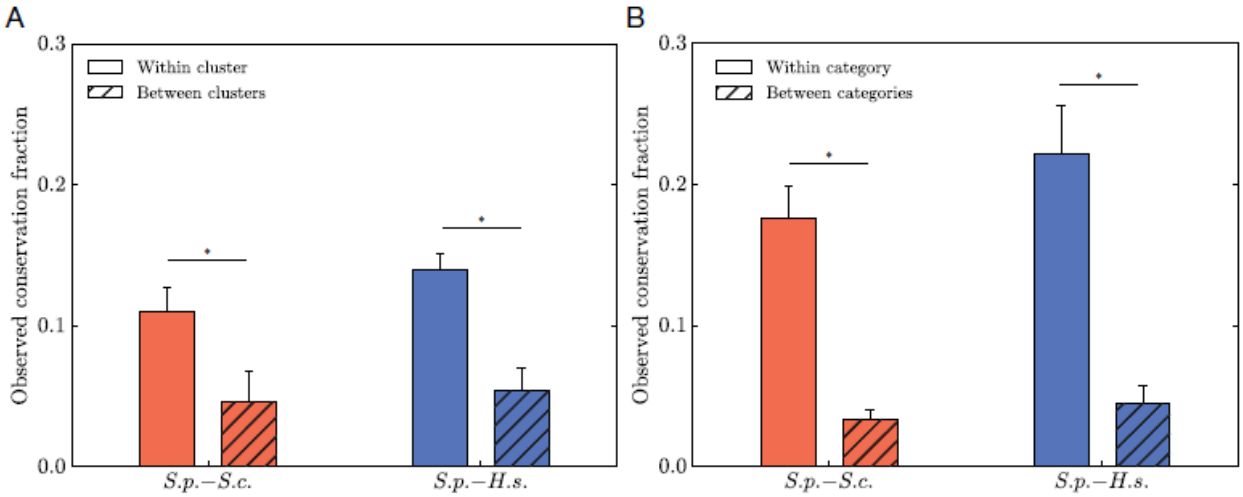


Figure 3.8 Interaction Conservation Within and Between Functional Modules

(A) Observed interaction conservation between proteins within or between topological clusters as computed by Affinity Propagation Clustering (APC). (B) Observed interaction conservation between proteins within or between GO slim categories as defined by PomBase. For both panels, conservation comparisons are between *S. pombe* and *S. cerevisiae* or human (*H. sapiens*). Data are shown as measurements + SE. * denotes significant (P<0.05); n.s. denotes not significant. Analyses by Jishnu Das.

we observed the same enrichment. Moreover, we find that the enrichment increases with the size of the complex, further supporting the notion that stoichiometric constraint influences the fate of WGD genes (Figure 3.10J).

Since WGD pairs are more functionally redundant than SSD pairs, these genes tend to be non-essential (Guan et al., 2007). It has also been shown that double deletions of these WGD pairs lead to a higher synthetic lethality rate than SSD pairs (Guan et al., 2007). Using a genome-scale genetic interaction map (Costanzo et al., 2010), we confirmed that deletion of WGD pairs is more likely to lead to synthetic lethality (Figure 3.9F; >6 fold difference in the fraction of synthetically lethal pairs, $P<10^{-10}$, *Z* test). Moreover, when we further stratify both groups of paralogs into pairs that are known to share interactors and pairs that have not been reported to share interactors, double deletions of the former are more likely to cause synthetic lethality than double deletions of the latter (Figure 3.9F; ~1.5 fold difference between the 2 sets for both SSD and WGD pairs, $P<0.05$ for both SSD and WGD pairs, *Z* test). This shows that paralog pairs that share interactors are more likely to be functionally redundant, regardless of whether they arose via SSD or WGD.

There have been conflicting reports in the literature regarding coexpression patterns of SSD and WGD pairs (Conant and Wolfe, 2006; Guan et al., 2007). Using a compendium of genome-wide expression datasets for *S. cerevisiae* genes (Yu et al., 2008), we found no significant difference in coexpression patterns of these pairs (Figure 3.10K). However, we find that SSD and WGD paralog pairs that share interactors are significantly more likely to be coexpressed than pairs that are not known to share interactors (Figures 3.9G and 3.10L-M; >10% more coexpressed for paralogs that are known to share interactors, $P<0.02$ in both cases, *Z* test). The tendency to be coexpressed among SSD pairs and WGD pairs that share interactors is the

same. Furthermore, among pairs that are not known to share interactors, WGD pairs tend to be more coexpressed than SSD pairs (Figures 3.9G and 3.10L-M; >10% more coexpressed for WGD paralogs compared to SSD paralogs, $P<0.02$ in all cases, Z test). These results show that for both duplication modes, because paralog pairs that are known to share interactors tend to be functionally redundant, the regulation of their gene expression also tends to be retained. Only for paralog pairs that are not known to share interactors is there a significant difference in coexpression between SSD and WGD paralogs, suggesting that even these WGD pairs might still be more functionally redundant than SSD pairs. It should be noted that, due to the incompleteness of current interactomes, paralog pairs could share interactors that are currently unreported.

The availability of proteome-wide interactomes helps dissect functional redundancy and divergence, and to some degree the regulation of expression, between paralogs. Overall, our results show that a WGD leads to greater functional redundancy while SSDs lead to greater functional diversification by sub-functionalization and neo-functionalization. Moreover, while there has been debate in the literature regarding the ubiquity of neo-functionalization (Gibson and Goldberg, 2009; He and Zhang, 2005), our results provide accurate measurements of the extent of neo-functionalization in the two yeasts.

3.5.6 Coevolution of conserved interactions revealed by cross-species interactome mapping

To further dissect the nature of conserved interactions, we implemented a cross-species interactome mapping approach to determine the prevalence of coevolution. We consider an interaction to be coevolved when its proteins have evolved in a coordinated manner to maintain

the interaction in different species, but have developed incompatible binding interfaces with orthologs of their partners. To determine whether conserved interactions are intact or coevolved, we test by Y2H whether a protein in one species can interact with the ortholog of its interacting partner in another species. If the cross-species interaction can occur, the interaction is intact (Figure 3.7C), otherwise it is coevolved between the two species (Figure 3.11A). For example, through our cross-species mapping, we discovered that interactions of farnesyltransferase subunit Cwp1 with other subunits Cpp1 and Cwg2 have coevolved between *S. pombe* and *S. cerevisiae*; Cwp1 cannot interact with either Ram1 or Cdc43, *S. cerevisiae* orthologs of Cpp1 and Cwg2, respectively (Figure 3.11B). A previous study showed that expression of Cwp1 cannot complement a non-functional mutant of its *S. cerevisiae* ortholog, Ram2 (Arellano et al., 1998). This suggests that Cwp1, although conserved between *S. pombe* and *S. cerevisiae* at the gene level, has evolved incompatible interaction interfaces with other farnesyltransferase subunits in *S. cerevisiae* and is thus unable to reconstitute an active enzyme complex.

It is known that evolution in protein folds is essentially the result of many random mutation events (Lockless and Ranganathan, 1999). However, since only a small fraction of changes that occur via random drift will satisfy the pairwise constraints necessary for interaction conservation, coevolution at the residue level only occurs at a few specific sites and is relatively rare (Talavera et al., 2015). Surprisingly we find that coevolution at the interaction level is not uncommon: ~33% and 50% of conserved interactions between *S. pombe* and *S. cerevisiae* or human are coevolved, respectively (Figure 3.11C). This shows that even among conserved interactions, only a few key alterations at important binding sites can make the cross-species interactions incompatible and the interactions coevolved. Thus, these sites are critical to protein binding and subsequent function, and changes at these sites alter protein interactions in a manner

Figure 3.9 Functional Divergence of Interactions Involving Paralogous Proteins

(A) Schematic representation of sub-functionalization and neo-functionalization. (B-C) Log odds ratios of sub-functionalization (B) for *S. pombe* and *S. cerevisiae* paralog pairs and (C) for *S. cerevisiae* SSD and WGD paralog pairs after correcting for divergence times. (D-E) Log odds ratios of neo-functionalization (D) for *S. pombe* and *S. cerevisiae* paralog pairs and (E) for *S. cerevisiae* SSD and WGD paralog pairs after correcting for divergence times. (F) Fraction of synthetic lethal pairs among SSD and WGD paralogs known or not known to share interactors. (G) Fraction of coexpressed pairs (*PCC*>0.4) among SSD and WGD paralogs known or not known to share interactors. Data are shown as measurements + SE. * denotes significant ($P<0.05$); n.s. denotes not significant. Analyses by Jishnu Das.

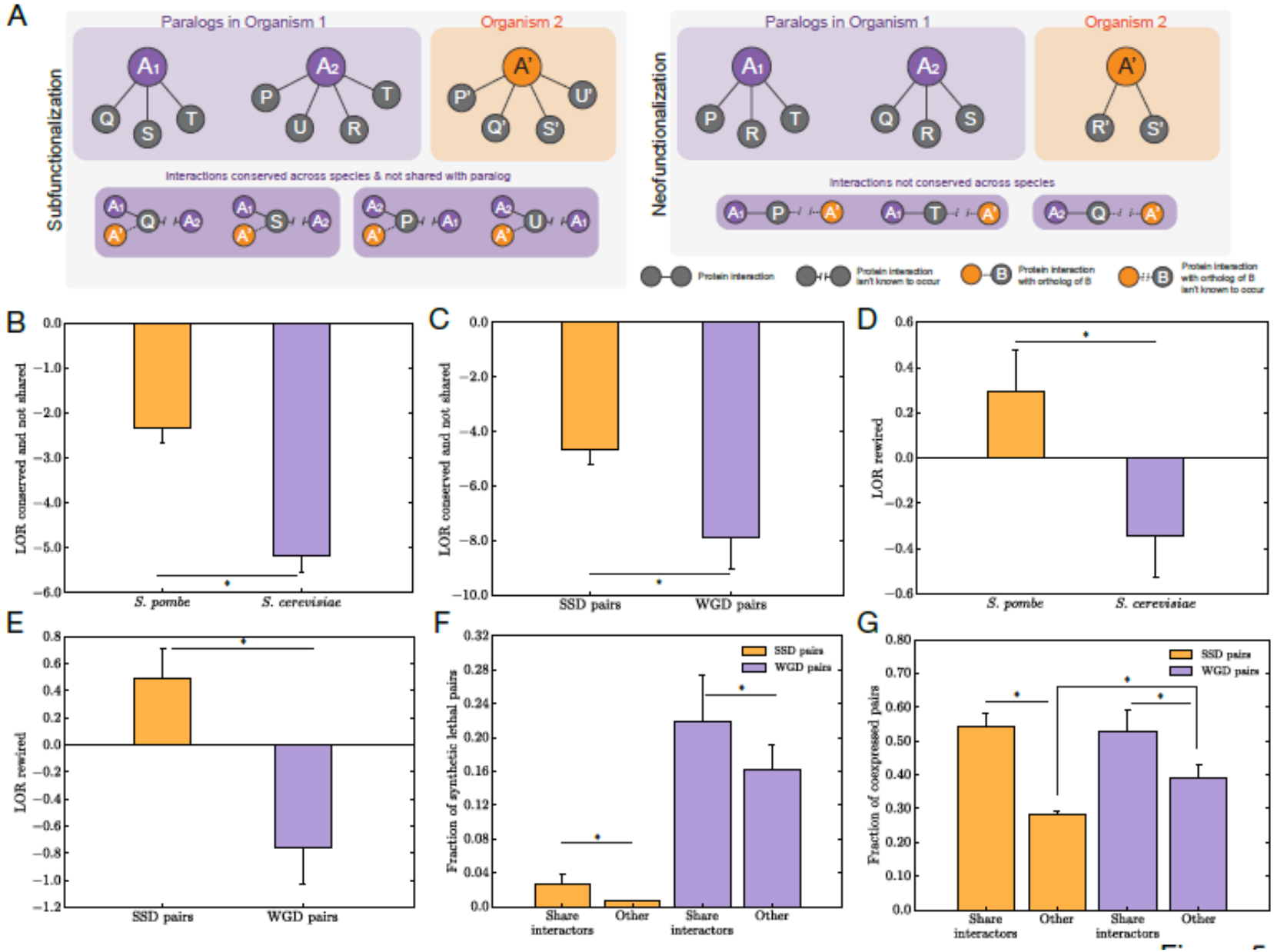
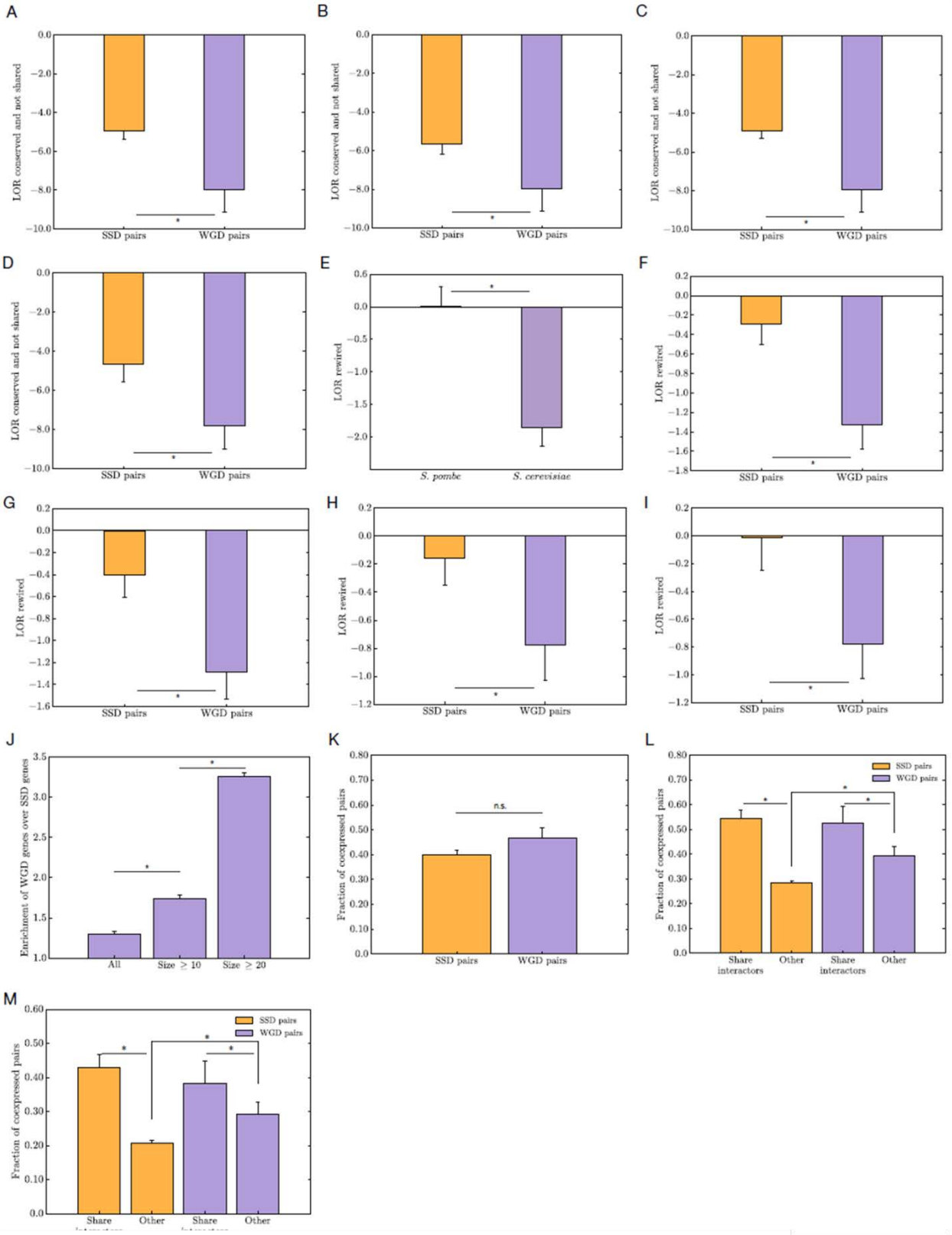

A
Subfunctionalization
Paralogs in Organism 1
Organism 2
Interactions conserved across species & not shared with paralog
Neofunctionalization
Paralogs in Organism 1
Organism 2
Interactions not conserved across species
Protein interaction
Protein interaction isn't known to occur
Protein interaction with ortholog of B
Protein interaction with ortholog of B isn't known to occur
B
LOR conserved and not shared
S. pombe
S. cerevisiae
C
LOR conserved and not shared
SSD pairs
WGD pairs
D
LOR rewired
S. pombe
S. cerevisiae
E
LOR rewired
SSD pairs
WGD pairs
F
Fraction of synthetic lethal pairs
SSD pairs
WGD pairs
Share interactors
Other
Share interactors
Other
G
Fraction of coexpressed pairs
SSD pairs
WGD pairs
Share interactors
Other
Share interactors
Other

Figure 3.10 Functional Divergence of Interactions Involving Paralogous Proteins

(A-D) Log odds ratio of sub-functionalization for *S. cerevisiae* SSD and WGD paralog pairs: (A) correcting for sequence evolution rates, (B) correcting for sequence identities, (C) without correcting for any covariates, and (D) where SSD and WGD pairs are defined using an independent dataset (Fares et al., 2013). (E) Log odds ratio of neo-functionalization for *S. pombe* and *S. cerevisiae* paralog pairs that do not share interactions. (F-I) Log odds ratio of neo-functionalization for *S. cerevisiae* SSD and WGD paralog pairs: (F) correcting for sequence evolution rates, (G) correcting for sequence identities, (H) without correcting for any covariates, and (I) where SSD and WGD pairs are defined using an independent dataset (Fares et al., 2013). (J) Enrichment of proteins from WGD paralog pairs compared to proteins from SSD paralog pairs in protein complexes of different sizes. (K) Fraction of SSD and WGD paralog pairs whose proteins are coexpressed (*PCC*>0.4), without separating pairs that are known to share and not known to share interactions. (L-M) Fraction of coexpressed pairs at other *PCC* cutoffs of (L) >0.3 or (M) >0.5 among SSD and WGD paralogs that are known to share and not known to share interactions. Data are shown as measurements + SE. * denotes significant ($P<0.05$); n.s. denotes not significant. Analyses by Jishnu Das.


A
B
C
D
E
F
G
H
I
J
K
L
M
LOR conserved and not shared
LOR rewired
SSD pairs
WGD pairs
S. pombe
S. cerevisiae
Enrichment of WGD genes over SSD genes
All
Size ≥ 10
Size ≥ 20
Fraction of coexpressed pairs
n.s.
Share interactors
Other

analogous to a single amino acid change disrupting protein interactions in human disease (Wang et al., 2012; Wei et al., 2014).

Among interactions for which we were able to determine coevolution status, we found that the likelihood for an interaction to be intact between *S. pombe*-*S. cerevisiae* and *S. pombe*-human is significantly higher than random expectation, while the likelihood for an interaction to be intact for one species pair and coevolved for the other species pair is significantly lower (Figure 3.11D; difference in log odds ratio=1.7, P=0.022, Z test). Thus, these intact interactions are likely involved in functions that have remained unchanged among yeasts and human throughout evolution.

We then investigated potential factors that could determine whether an interaction is intact or coevolved with respect to another species. We find that overall sequences of proteins involved in intact interactions tend to be better conserved across species than sequences of proteins in coevolved interactions (Figure 3.11E; 18.0% higher, $P=2.1\times10^{-4}$ for *S. pombe*-human, U test). High sequence conservation may indicate higher levels of evolutionary constraint existing within the local network neighborhood of a given interaction. In fact, we find that proteins involved only in intact interactions have twice the number of interactors as compared to proteins involved in only coevolved interactions (Figure 3.11F; $P=1.1\times10^{-3}$, U test), suggesting that the added evolutionary constraint of maintaining many interacting partners may prevent the coevolution of two interacting proteins. Finally, we find that the most highly evolutionarily correlated inter-protein residue pairs in coevolved interactions are significantly more correlated than top residue pairs in intact interactions (Figure 3.12A; $P<10^{-10}$, U test), suggesting that the maintenance of coevolved interactions involves compensatory changes at the amino acid residue level.

3.4.7 Implications of FissionNet for the study of human disease

We explored the relevance of FissionNet to human disease by considering the context of known human disease mutations from HGMD (Stenson et al., 2014) within proteins of the human interactome conserved in *S. pombe*. We find that among human interactions conserved in either *S. pombe*, *S. cerevisiae*, or both, ~40% of inter-protein pairs of disease mutations cause the same disease (Figure 3.13A). This is significantly higher than in human interactions that are not reported to be conserved in either yeast or cannot be conserved in either due to lack of protein orthologs (Figure 3.13A; $P<10^{-10}$ for all pairwise comparisons, *Z* test). Based on these results, mutations that break specific protein-protein interactions to cause diseases may be overrepresented among interactions conserved in model organisms. From a global network view, FissionNet may be highly relevant to the study of human disease based on the large portion of *S. pombe* interactions in which both proteins have human orthologs with known germline disease or somatic cancer-associated mutations (Figure 3.13B; 902 interactions) (Forbes et al., 2015; Stenson et al., 2014).

To demonstrate the plausibility of studying specific human disease mutations using FissionNet, we explored whether human disease mutations that disrupt human interactions intact in *S. pombe* also disrupt the corresponding interactions of the fission yeast orthologs. We focused on three examples: two Mendelian disease variants (Stenson et al., 2014) that disrupt the human NMNAT1-NMNAT1 and PCBD1-PCBD1 interactions and one population variant from the Exome Sequencing project (Fu et al., 2013) that disrupts the human SNW1-PPIL1 interaction. We find that introducing these human protein residue changes into their *S. pombe* orthologs also disrupts the fission yeast interactions (Figure 3.13C). These results indicate that cross-species interactome mapping enables investigation of whether interaction interfaces are altered at the

molecular level between model organisms and human, a finding with potentially far-reaching implications for the study of protein function and human disease.

Our results regarding gene duplication modes may also be relevant to the study of human disease. We find that human WGD paralog pairs have a significantly higher likelihood to be involved in the same disease compared to human SSD paralog pairs, in agreement with our observation that WGD paralog pairs tend to be functionally redundant (Figure 3.13D; 7 fold difference in the fraction of WGD and SSD pairs that cause the same disease, $P<10^{-10}$, Z test). Thus, our findings have direct implications for understanding the functional roles of paralogous genes, from yeasts to human.

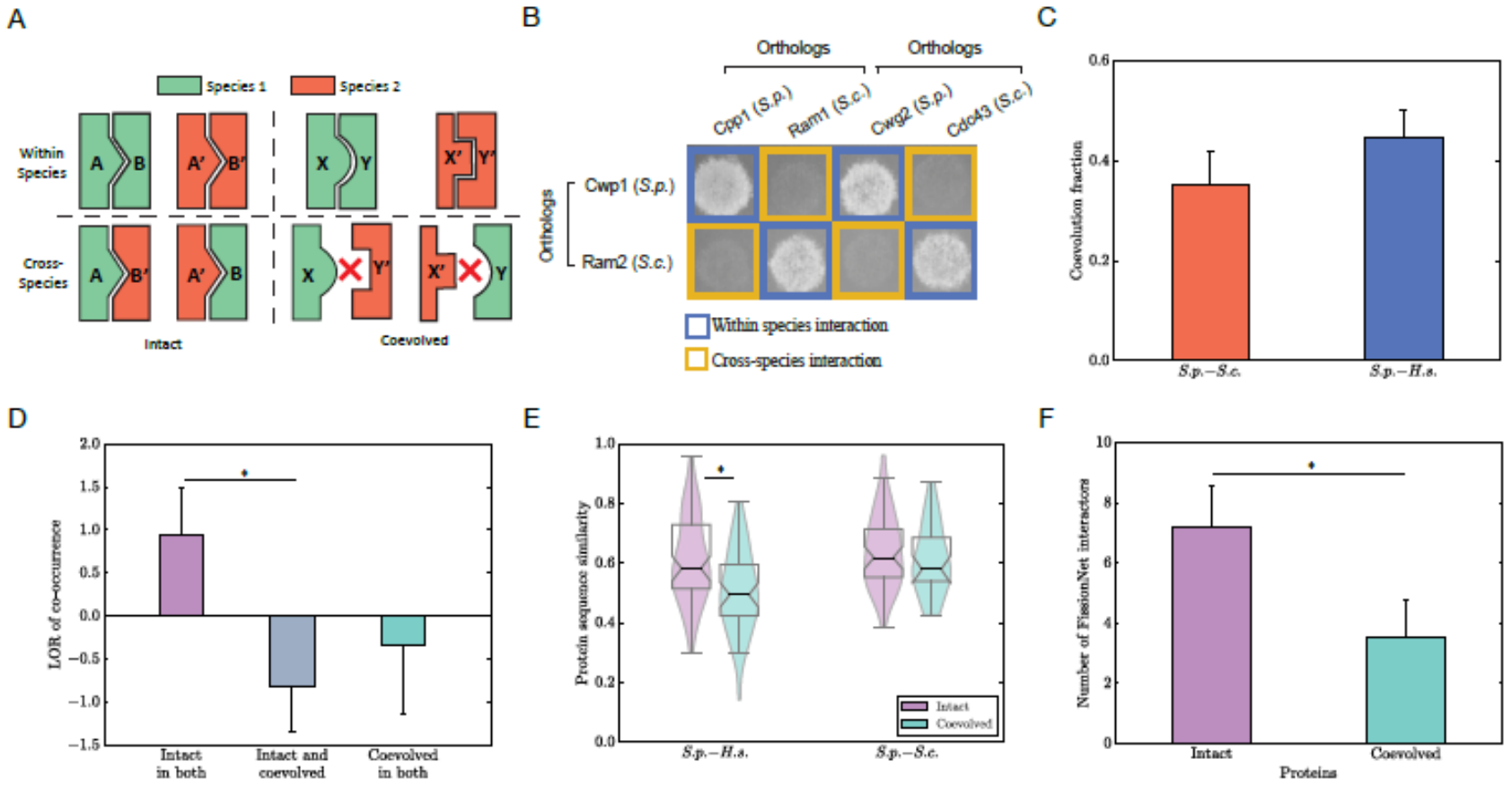
3.6 CONCLUSIONS AND DISCUSSION

FissionNet provides a wealth of functional information. For example, we find that the Atf1-Cid12 interaction mediates silencing at Atf1-target genes *hsp16* and *hsp104*. It has been shown that the RNAi pathway is involved in silencing of these genes (Woolcock et al., 2012). Hence, it is possible that the Atf1-Cid12 interaction is part of an RNAi-dependent regulatory pathway.

By comparing FissionNet to protein networks in budding yeast and human, we have shown that the molecular bases for interaction conservation among orthologous proteins are complex and different from those that underlie gene conservation. This is highly relevant to the use of the two yeasts as model organisms as there are functions that can be better studied using fission yeast. We find that divergence across species is not completely dictated by sequence level changes, suggesting that rewiring of interactomes plays an important role in species evolution. Additionally, our finding that proteins in a significant fraction of conserved interactions have

Figure 3.11 Intact and Coevolved Interactions

(A) Schematic representation of conserved protein interactions that are either intact or coevolved. (B) Within- and cross-species Y2H detects coevolved interactions. (C) Fraction of *S.p.* interactions that are coevolved with respect to *S.c.* or human (*H.s.*). (D) Log odds ratio of co-occurrence of intact and coevolved interactions between *S.p.-S.c.* and *S.p.-H.s.* (E) Overall protein sequence similarity of *S.p.* proteins involved in intact or coevolved interactions. (F) Number of interactors for proteins involved in intact or coevolved interactions. Data are shown as measurements + SE. * denotes significant ($P<0.05$); n.s. denotes not significant. Panels C, D, E, and F by Jishnu Das and Michael Meyer.


A
Species 1
Species 2
Within Species
Cross-Species
A
B
A′
B′
X
Y
X′
Y′
Intact
Coevolved
B
Orthologs
Orthologs
Cpp1 (S.p.)
Ram1 (S.c.)
Cwg2 (S.p.)
Cdc43 (S.c.)
Cwp1 (S.p.)
Ram2 (S.c.)
Orthologs
Within species interaction
Cross-species interaction
C
Coevolution fraction
0.0
0.2
0.4
0.6
S.p.–S.c.
S.p.–H.s.
D
LOR of co-occurrence
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
*
Intact in both
Intact and coevolved
Coevolved in both
E
Protein sequence similarity
1.0
0.8
0.6
0.4
0.2
0.0
*
Intact
Coevolved
S.p.–H.s.
S.p.–S.c.
F
Number of FissionNet interactions
10
8
6
4
2
0
*
Intact
Coevolved
Proteins

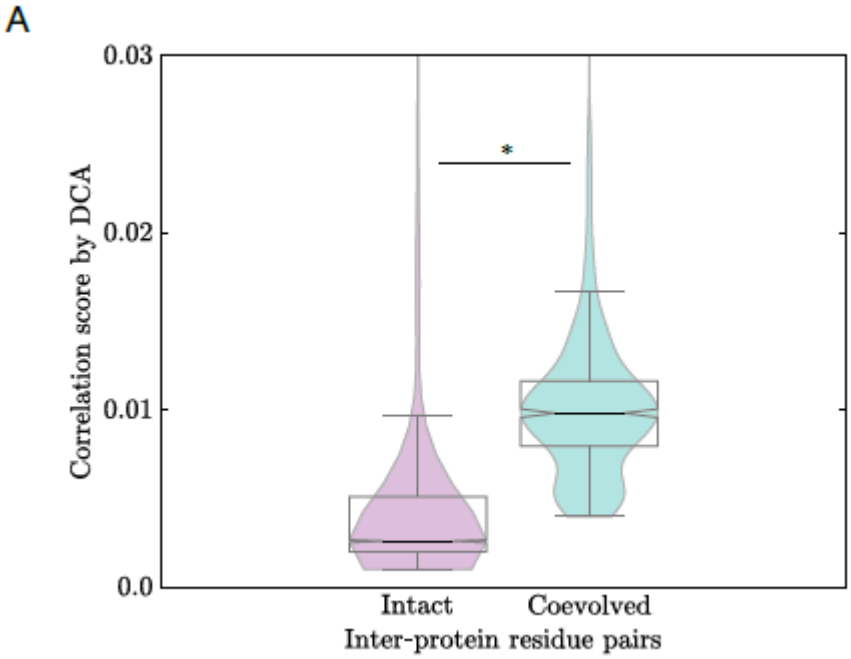


Figure 3.12 Inter-protein Residue Pairs in Coevolved Interactions are More Correlated than in Intact Interactions

(A) Co-evolution analysis of proteins involved in intact or co-evolved interactions using DCA. * denotes significant (P<0.05). Analysis by Michael Meyer.

undergone coevolution to maintain interactions has major implications for studies reliant on the expression of human proteins in model organisms to identify functional mechanisms (Tardiff et al., 2013).

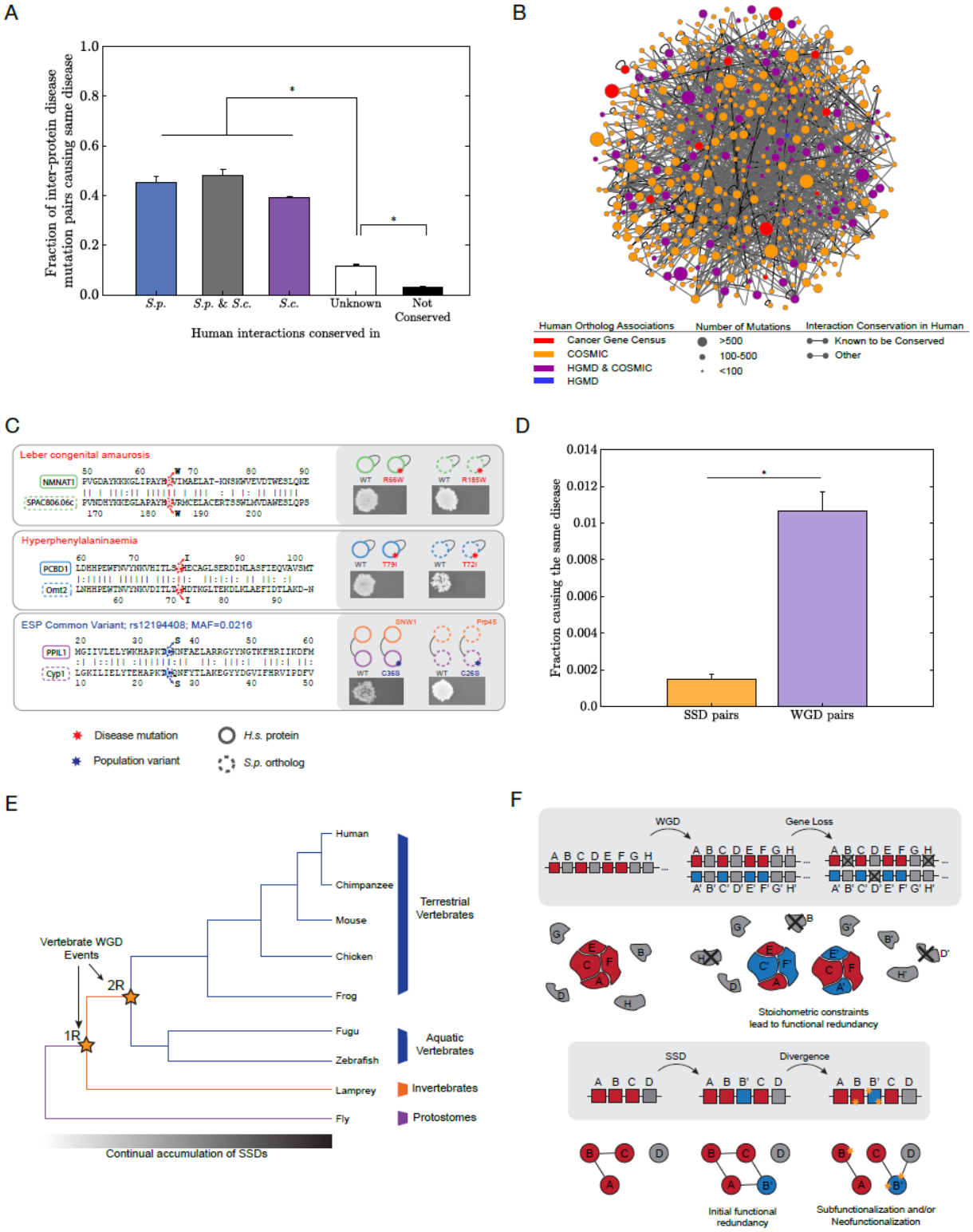
Gene duplication is a key process shaping evolution (Figure 3.13E). Our results show that paralogs arising via WGD are under strong constraints to maintain stoichiometric ratios with their interaction partners and, hence, tend to maintain functional redundancy; on the other hand, duplicates arising via SSDs are not under such strong constraints and are more likely to gain novel functions (Figure 3.13F). For example, it has previously been reported that duplicate copies of the *SRGAP2* gene that arose via segmental duplications (SSD-like events) have gained new functions related to brain development specifically in the human lineage (Dennis et al., 2012).

Gene duplications play an important role in the evolutionary mechanism governing speciation as well as the evolution of developmental and morphological complexity in vertebrates (Rensing, 2014; Ting et al., 2004). For example, two rounds of WGD have been predicted in the origin of jawed vertebrates (Figure 3.13E) (Kasahara, 2007). During speciation, while certain key functions need to be evolutionarily preserved, new functions are necessary for differential adaptation between species (Ting et al., 2004). Previous studies have identified how duplication events can lead to functional changes through gene dosage alterations (Papp et al., 2003). Our results help establish on a proteomic scale that paralogs arising via WGD are more likely to preserve functions and provide robustness for important cellular functions, while paralogs arising via SSDs are more likely to contribute to novel functions gained by specific species. These findings further our understanding of human biology and disease.

Our analyses focus on budding yeast, fission yeast, and human, as they are the only three

Figure 3.13 FissionNet as a Resource for Studying Human Disease

(A) Fraction of inter-protein HGMD mutation pairs that cause the same disease in human interactions with regard to their conservation status in *S. pombe* and *S. cerevisiae*. (B) Largest connected subcomponent of FissionNet wherein all proteins have human orthologs with known germline disease or somatic cancer-associated mutations. (C) Impact of human disease mutations and a population variant on intact interactions between human and fission yeast. (D) Fraction of human SSD and WGD paralogs that cause the same disease. (E) The 2R hypothesis predicts two recent WGD events leading to the vertebrate lineage. (F) WGD can lead to more functional redundancy through targeted gene loss that maintains stoichiometric ratios of protein products. SSD leads to more neo-functionalization and sub-functionalization through alterations to initially redundant paralogs. Data are shown as measurements + SE. * denotes significant (P<0.05); n.s. denotes not significant. Panel A analysis by Michael Meyer. Panel D analysis by Jishnu Das.


A
Fraction of inter-protein disease mutation pairs causing same disease
1.0
0.8
0.6
0.4
0.2
0.0
*
*
S.p.
S.p. & S.c.
S.c.
Unknown
Not Conserved
Human interactions conserved in
B
Human Ortholog Associations
Cancer Gene Census
COSMIC
HGMD & COSMIC
HGMD
Number of Mutations
>500
100-500
<100
Interaction Conservation in Human
Known to be Conserved
Other
C
Leber congenital amaurosis
NMNAT1
SPAC806.06c
WT
R56W
R185W
Hyperphenylalaninaemia
PCBD1
Omt2
T79I
T72I
ESP Common Variant; rs12194408; MAF=0.0216
PPIL1
Cyp1
SNW1
Prp45
C36S
C26S
Disease mutation
Population variant
H.s. protein
S.p. ortholog
D
Fraction causing the same disease
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0.0
*
SSD pairs
WGD pairs
E
Human
Chimpanzee
Mouse
Chicken
Frog
Fugu
Zebrafish
Lamprey
Fly
Terrestrial Vertebrates
Aquatic Vertebrates
Invertebrates
Protostomes
Vertebrate WGD Events
2R
1R
Continual accumulation of SSDs
F
WGD
Gene Loss
A B C D E F G H
A' B' C' D' E' F' G' H'
Stoichometric constraints lead to functional redundancy
SSD
Divergence
A B C D
A B B' C D
Initial functional redundancy
Subfunctionalization and/or Neofunctionalization

eukaryotic organisms for which we have proteome-scale interactome networks using our version of the Y2H assay (>50% of all protein pairs screened). Once more interactome networks are systematically generated in other species, using assays with measured sensitivity and specificity, the comparative network analysis framework established in this study can be readily applied to further elucidate the extent and nature of the evolution of protein functions across many species.

3.7 ACKNOWLEDGEMENTS

We thank D. Moazed for *S. pombe* strains provided for the analyses of centromeric silencing; S. Forsburg, H. Madhani, and J. Nakayama for providing plasmids for protein expression in *S. pombe*. We are grateful to M. Smolka for experimental advice; D. Barbash, A. Clark, S. Emr, S. Grewal, D. Lin, D. Shalloway, and members of the Yu lab for discussions. All DNA microarray data are available through NCBI Gene Expression Omnibus (GSE74354). This work was supported by Marie Curie Cancer Care, Association for International Cancer Research and Cancer Research UK to M. Trickey and H. Yamano., NHGRI grants HG001715 and HG004233, Krembil Foundation and Canada Excellence Research Chairs Program to F. P. Roth, Canada Research Chairs Program and grants NSERC RGPIN-2014-03892, NSF CCF-1219007, and CFI JELF-33732 to Y. Xia, NIGMS grants GM098634 to J. A. Pleiss, and GM097358 and GM104424 to H. Yu.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

The protein interactome of organisms underlie a major portion of their functional landscape. In this thesis, I present a proteome-wide interactome network for the fission yeast, *S. pombe*, which we named FissionNet. This network represents only the second reported proteome-wide protein interactome network for any eukaryote in which >95% of the proteome-space have been systematically and experimentally queried for binary protein-protein interactions (the first reported "complete" interactome network belonging to the budding yeast, *S. cerevisiae*) (Yu et al., 2008). It is important to note that StressNet represents a subset of FissionNet.

We ensured the high quality of FissionNet using several strict filtering and validation methods. During the high-throughput experimental pipeline, only interactions that tested positive in our Y2H assays three out of three times per screen were considered true Y2H positives (controlling for biological bias). I also tested for, and removed, all autoactivating protein constructs which would have resulted in Y2H false positives. We also validated FissionNet using an orthogonal assay called PCA (PCA experiments performed by Nurten Akturk). We found that the overall PCA detection rate of a random subset of FissionNet interactions was the same as the detection rate of a positive reference set using the same assay (Chapter 3, Figure 3.1B; P=0.34, Z test). For validation of FissionNet interactions, Jishnu Das and Michael Meyer showed that, overall, protein pairs annotated to interact in FissionNet are more likely to be colocalized and to share similar biological functions as compared to random pairs of *S. pombe* proteins. Moreover, gene pairs represented by FissionNet interactions are more correlated in terms of gene expression

compared for random *S. pombe* gene pairs. These experimental protocols and analyses maintain that FissionNet interactions are of high quality and that they are supported by orthogonal biological datasets (PCA interactions, protein colocalizations, etc).

I illustrated the resourcefulness of FissionNet by using it as a launching point to experimentally and empirically make a number of biological discoveries. First, I discovered a novel interaction between Snr1 and the well-characterized stress-response kinase Sty1 from my Y2H screens (Chapter 2). I showed that Snr1 is also involved in stress-response; loss of *snr1* results in hypersensitivity of *S. pombe* to a number of stressors such as osmotic stress.

Then, I discovered new gene regulatory functions for several additional proteins (Chapter 3). I found that the casein kinase Hhp1 is a novel interactor of the RNAi-induced transcriptional silencing (RITS) complex core protein, Tas3. I found that *hhp1Δ* cells lose the ability to properly silence centromeric transcription and also lose the heterochromatin marker, H3K9me2. Tas3, along with RITS complex components Ago1 and Chp1, is known to bind to chromatin regions such as the centromere and to recruit heterochromatin modifiers such as the sole *S. pombe* lysine-4 methyltransferase Clr4 (Verdel et al., 2004). The entire RITS complex recognizes target chromatin regions by first binding to RNAi whose sequence are complementary to that of the target chromatin. On the other hand, Hhp1 is a casein kinase that mediates serine/threonine phosphorylation. It has known roles in regulating cytokinesis and meiosis in fission yeast (Johnson et al., 2013; Petronczki et al., 2006). In response to stress cues during mitosis, Hhp1 will initiate the Sid4-Dma1 pathway through phosphorylation of Sid4 to induce the arrest of cytokinesis and, thus, prevent full cell division to mitigrate damage arising from aberrant mitosis (Johnson et al., 2013). Moreover, it has been observed that Hhp1 can change localization between the nucleus and cytosol, which strongly suggests that this protein may play a role as a

"middle man" to regulate the interplay of multiple processes within the cell (Matsuyama et al., 2006). Since, proper centromeric silencing is vital to proper nuclear division during mitosis (centromeric silencing is a prerequisite to proper kinetochore assemble at the centromere) (Pidoux et al., 2003) and mass spectrometry experiments have shown the presence of phosphorylated serine residues on Tas3 (Wilson-Grady et al., 2008), it might be that Hhp1 must phosphorylate Tas3 inside the nucleus to induce centromeric silencing. This could be through a variety of different mechanisms, such as allowing RITS complex to fully form, allowing RITS complex to bind to their target chromatin, and/or allowing RITS complex to recruit downstream heterochromatin proteins. Aside from heterochromatin initiation, another possibility is that Hhp1 might play a crucial role in the maintenance of heterochromatin. This possibility might not be likely because mass spectrometry using Hhp1 as bait have previously failed to reveal Tas3 as an interactor (Johnson et al., 2013). This suggests that the interaction might be transient and condition-specific. Either way, if some kind of unknown mitotic stress does arise, Hhp1 might translocate away from the centromere/nucleus to elicit signals to prevent completion of cell division altogether.

Other discoveries were regarding the roles of the stress-response transcription factor Atf1 and the poly-A polymerase Cid12 on transcriptional regulation of stress response genes (Chapter 3 and Appendix I). Specifically, I showed that loss of Cid12 leads to upregulation of Atf1-target genes, *hsp16* and *hsp104* (target genes implying that their promoters are bound by Atf1). Moreover, I found that disrupting the interaction of Cid12 and Atf1, without losing the proteins or affecting centromeric silencing, also cause *hsp16* and *hsp104* upregulation. This indicates that the Cid12-Atf1 interaction mediates transcriptional repression of the genes. Follow-up experiments (Appendix I) show that loss of RNAi and centromere silencing factors, including

Tas3 and Hhp1, also abrogate *hsp16* silencing. Moreover, the upregulation appears equivalent to that normally induced during heat-stress, suggesting that these proteins may be sufficient to completely repress transcription under non-stressed conditions. Surprisingly, I found that loss of Atf1 alone was insufficient to remove *hsp16* repression. Moreover, loss of Clr4 alone only slightly caused derepression. However, double mutants *atf1Δ clr4Δ* caused dramatic depression, suggesting that Atf1 and Clr4 affect repression in parallel pathways. Importantly, this suggests that Cid12 might actually function in both Atf1 and Clr4 pathways to mediate *hsp16* silencing. This represents the exciting possibility that Atf1 functions within the RNAi pathway, in parallel with the H4 methylation pathway, to suppress aberrant gene expression. This would be highly surprising since it has been previously shown that Atf1 works in a parallel pathway with RNAi to silence the mating-type locus (Jia et al., 2004). The broader implications of such a finding would be that different genomic regions could leverage the same protein factors in different ways to perform similar tasks. To address this question, whole-genome experiments must be combined with classical genetics to begin dissecting the interplay of various silencing pathways.

In collaboration with Jishnu Das and Michael Meyer (for all computational analyses), we find that, throughout evolutionary time, there have been much less conservation of interactions between *S. pombe* and *S. cerevisiae* than would be expected simply based on protein orthology. In other words, we estimate that for only ~40% of FissionNet interactions are their orthologous budding yeast interactions conserved. However, if we assumed *a priori* that a pair of proteins with orthologs in both yeasts should intereact if they interact in any one species, then the expected conservation rate should be ~64% (based on the fact that ~80% of *S. pombe* proteins have at least one ortholog in *S. cerevisiae*). Surprisingly, we found that FissionNet interactions are much more conserved in human (~60% of FissionNet interactions conserved in human).

Moreover, we find that the conservation rates are not evenly distributed across biological processes. For example, we find that FissionNet interactions are more conserved in human than in *S. cerevisiae* for processes such as chromosomal segregation and cell cycle regulation whereas there was not a significant difference for processes such as transport or nitrogen metabolism. This corroborates with the traditional history of *S. pombe* as being the model yeast for studying cell cycle and chromosomal dynamics (Wood et al., 2002).

To further dissect the nature of protein interaction evolution, we measured the fraction of intact versus coevolved conserved interactions. We find that, when comparing *S. pombe* to *S. cerevisiae* or human, coevolution is quite prevalent (~60%). Coevolution of interacting protein pairs correlate with factors such as intact-coevolved status of local interactions or number of interactions. This suggests that, beyond the simplistic explanation of organismal fitness, protein interactome network restraints most likely also play key roles in shaping the particular modes of functional evolution. Moreover, we might be able to leverage these results to help us better understand mechanisms of human diseases in yeast. I find two cases (Chapter 3) where a disease-associated human mutation can disrupt a conserved and intact interaction in human and *S. pombe*. Since fission yeast is a much more tractable model organism, it would be possible to further study the molecular impact of such disease-associated mutations in the yeast. For example, if the conserved interactions are part of a broader conserved pathway, it would be possible to assess the impact on this pathway of the mutation-of-interest. However, this would be impossible for conserved but coevolved interactions because, by definition, mutations cannot be mapped across species.

In conclusion, my thesis covers a major resource for the scientific community (FissionNet) and I explore various uses of the network to uncover new biology. I surmise that

experts in various biological processes would be able to make better use of FissionNet than I can through expertise insights. Together, my work contributes to the larger systems biology goal of being able to accurately see the entire cell as a massive and deconvoluted system of networks.

APPENDIX I

ADDITIONAL RNAI AND HETEROCHROMATIN FACTORS ARE INVOLVED IN TRANSCRIPTIONAL SILENCING OF HEATSHOCK GENES

SUMMARY

As detailed in Chapter 3, I discovered that Cid12 is essential to transcriptionally repress heatshock genes in *S. pombe*. In this Appendix I, I detail follow-up experiments that I performed to further explore possible mechanisms and additional factors that regulate the heatshock gene *hsp16*. I find that RNAi proteins play a dominant role in *hsp16* repression under non-stressed conditions. I also find that members of the RITS complex, Tas3 and Chp1, play a role in repression. Surprisingly, loss of *atf1* or the heterochromatin factor *clr4* does not cause major derepression but *hsp16* is overexpressed when both components are deleted in *S. pombe*. These data strongly suggest that Atf1 and Clr4 are likely in parallel pathways that repress *hsp16* expression and, given results presented in Chapter 3, Cid12 is likely to be involved in both pathways.

CONTRIBUTIONS

Haiyuan Yu and I designed and interpreted all experiments. I performed all experiments.

INTRODUCTION

The poly-A polymerase Cid12 in *S. pombe* has a well-characterized role in centromeric silencing. There, Cid12 is part of the RNA-directed RNA polymerase complex (RDRC) with the RNA-directed RNA polymerase, Rdp1, and an RNA helicase, Hrr1 (Motamedi et al., 2004). The RDRC is responsible for generating dsRNA using centromeric ssRNA as template. Moreover, the RDRC is targeted to the centromere through interaction with the RITS complex, which is bound through Chp1-chromatin binding. The synthesized dsRNA is loaded onto the RNAi protein, Dcr1, which cleaves to generate siRNA that can bound by Ago1. Recall that Ago1 is a member of the RITS complex. Ago1, in complex with Chp1 and Tas3 and pre-loaded with centromeric siRNA, targets the entire RITS complex to the centromere through DNA-RNA strand complementarity. Essentially, the RDRC and RITS complex form a positive-feedback loop that assembles and maintains both complexes at centromeric loci and are responsible for nucleating the formation of centromeric heterochromatin. Formation of heterochromatin effectively causes centromeric transcriptional silencing (Verdel et al., 2004).

Interestingly, I discovered that Cid12 is also required for transcriptional silencing of the Atf1-target heatshock gene *hsp16* (Chapter 3). Atf1 has been shown, through ChIP-chip experiments, to bind the promoter of *hsp16* which presumably suggests some kind of gene regulation (Eshaghi et al., 2010). Loss of the RNAi factor Dcr1 has also been shown to lead to depression of *hsp16* (Woolcock et al., 2012). Thus, I explored the possibility that the same factors involved in centromeric silencing also repress *hsp16* expression. Since Atf1 does not bind to the centromere, the central question is how different genomic regions can leverage a common set of proteins to perform similar functions.

MATERIALS AND METHODS

S. pombe culturing

S. pombe wild-type strains were cultured in yeast extract with supplements (YES) (5g/L yeast extract, 3% glucose, 225mg/L of adenine, histidine, leucine, uracil and lysine hydrochloride). Deletion strains with the *kanMX* cassette were selected for on YES plates with 150mg/L G418 (Calbiochem). Deletion strains with the *natMX* cassette were selected for on YES plates with 100mg/L noursethecin/LEXSY NTC (Mitegen). For the temperature-sensitivity cell growth assay on plate, wild-type and mutant strains were grown on YES plate without drug selection. For heatshock experiments using liquid cultures, cells were cultured in YES at 30°C until log-phase. After, cultures were shifted to 37°C to induce heatshock for variable durations prior to harvesting. Non-stressed condition entails culturing of cells at 30°C.

Gene deletion in *S. pombe*

Genes were deleted using a PCR homology-based approach. Briefly, the pFA6a-KanMX6 module was used to replace the entire ORF-of-interest by generating, through stitch PCR, the cassette flanked by ~300bp sequences homologous to 300bp upstream and downstream of the ORF in the genome. The resultant PCR product was purified using Microelute Cycle-Pure kit (Omega) and transformed into *S. pombe* cells using a lithium acetate approach. Primers used for generating the deletions were listed in Chapter 3 (Table 3.2) or in Table Appendix I.1.

Reverse transcription PCR (RT-PCR)

RNA isolation procedures were performed as described in (Inada and Pleiss, 2010) and detailed in Chapter 3 Materials and Methods.

Primer	Sequence (5' → 3')	Purpose
Clr4-Del-UP-fwd	CATGAATTGGCATTGCTAGTTTGG	Delete Clr4
Clr4-Del-UP-rev	ttaattaacccggggatccgTCGCAAAACTAATAACCTCTTGTTTG	Delete Clr4
Clr4-Del-pFA-fwd	CAAACAAGAGGTTATTAGTTTTGCGACGGATCCCCGGGTTAATTAA	Delete Clr4
Clr4-Del-pFA-rev	TTGGAGTCAACCAGTAATAAATTAGCGAATTCGAGCTCGTTTAAAC	Delete Clr4
Clr4-Del-DN-fwd	gtttaaacgagctcgaattcGCTAATTTATTACTGGTTGACTCCAA	Delete Clr4
Clr4-Del-DN-rev	CTTCTGAAGGACGGTCCATTAC	Delete Clr4
Swi6-Del-UP-fwd	AATCACTAGGAAATTGAGATGCTT	Delete Clr4
Swi6-Del-UP-rev	ttaattaacccggggatccgTTTTCACTTGTCTTAATATGAAAATAAACG	Delete Swi6
Swi6-Del-pFA-fwd	CGTTTATTTTCATATTAAGACAAGTGAAAACGGATCCCCGGGTTAATTAA	Delete Swi6
Swi6-Del-pFA-rev	AGAATTTTTTAAAGGAACACAAAAAAAGAATTCGAGCTCGTTTAAAC	Delete Swi6
Swi6-Del-DN-fwd	gtttaaacgagctcgaattcTTTTTTTGTGTTCCTTTAAAAAATTCT	Delete Swi6
Swi6-Del-DN-rev	GAATGTCAAACCATAACGACA	Delete Swi6

Appendix Table I.1 Primers Used in this Study

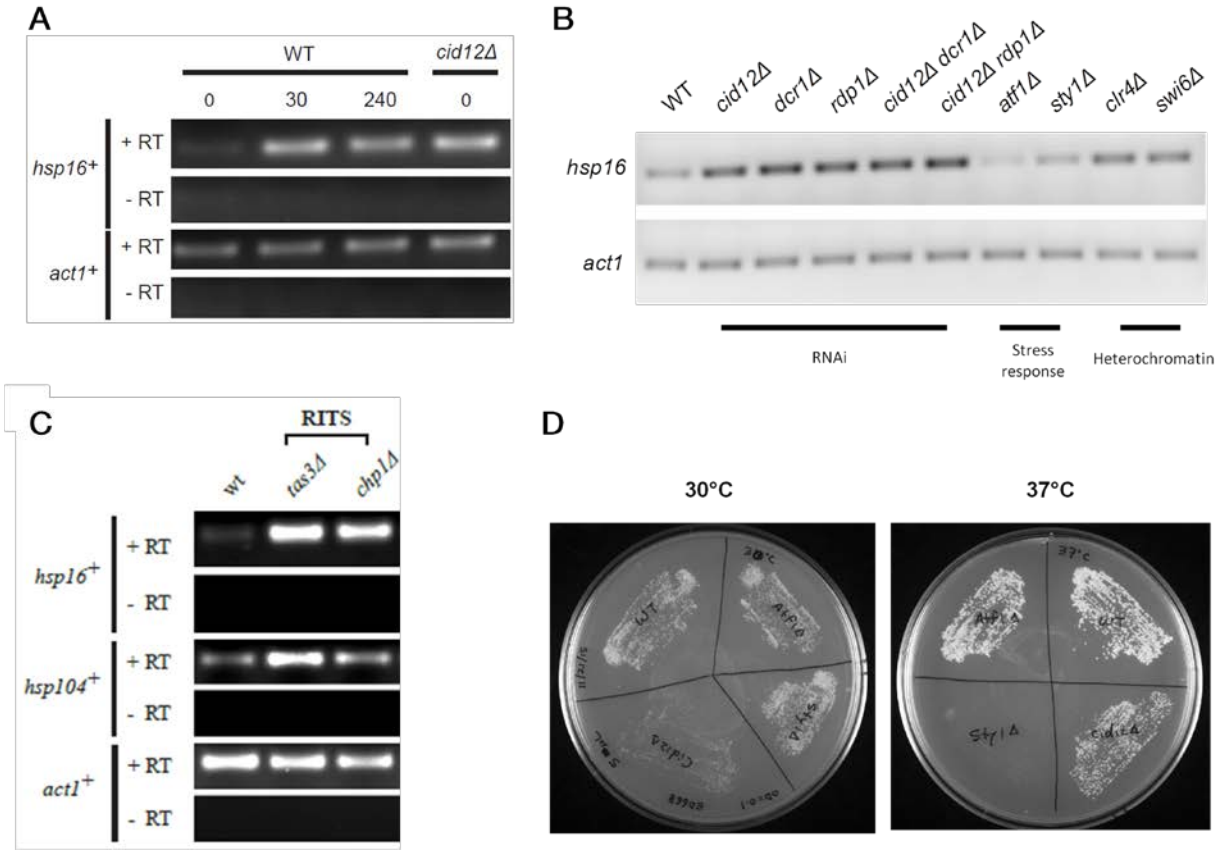
Primer	Sequence (5' → 3')	Purpose
Rdp1-UP-fwd	CTATACGTAAAGGAGCCTTCAC	Delete Rdp1
Rdp1-UP-rev	ttaattaacccggggatccgTGCTTTACATTGCCATACAAAATC	Delete Rdp1
Rdp1-pFA6a-fwd	GATTTTGTATGGCAATGTAAAGCACGGATCCCCGGGTTAATTAA	Delete Rdp1
Rdp1-pFA6a-rev	AAGCTTCGAATGTATATTCGATGAATTCGAGCTCGTTTAAAC	Delete Rdp1
Rdp1-DN-fwd	gtttaaacgagctcgaattcATCGAATATACATTCGAAGCTT	Delete Rdp1
Rdp1-DN-rev	GTGAAAATTAGATTAAGCTTGGCAG	Delete Rdp1
Atf1-UP-fwd	ATCTCCTTAATCTCCTCCTC	Delete Atf1
Atf1-UP-rev	ttaattaacccggggatccgAATTGAAGAATTTATGCTTTAACACT	Delete Atf1
Atf1-pFA6a-fwd	AGTGTTAAAGCATAAATTCTTCAATTCGGATCCCCGGGTTAATTAA	Delete Atf1
Atf1-pFA6a-rev	AGACCTTTTCAGATCAAAAACAGTGAATTCGAGCTCGTTTAAAC	Delete Atf1
Atf1-DN-fwd	gtttaaacgagctcgaattcACTGTTTTTGATCTGAAAAGGTCT	Delete Atf1
Atf1-DN-rev	TGTTGTCCAAACGATATTATAGTGA	Delete Atf1
Sty1Del-UP_fwd	TACAAGCAAACACCACAATC	Delete Sty1
Sty1Del-UP_rev	TTAATTAACCCGGGGATCCGTTTATTCAAACTGGTTACAAAAAGGAC	Delete Sty1
PFA6a-Sty1_fwd	TTGTAACCAGTTTGAATAAACGGATCCCCGGGTTAATTAA	Delete Sty1
PFA6a-Sty1_rev	AGGCTTTATCTACAACTTGTGAATTCGAGCTCGTTTAAAC	Delete Sty1
Sty1Del-DN_fwd	GTTTAAACGAGCTCGAATTCACAAGTTGTAGATAAAGCCTTAAAAGTTGTTC	Delete Sty1
Sty1Del-DN_rev	ACACCACACTTGAAAATCGC	Delete Sty1

(Appendix Table I.1 continued)

RESULTS AND CONCLUSIONS

Although I have shown previously that, under non-stressed conditions, *hsp16* is overexpressed in *cid12Δ* cells, it was not clear whether Cid12 played a major or minor role in silencing. I induced temperature stress by exposing wild-type *S. pombe* cells to 37°C culturing condition for 0, 30, or 240 minutes. I found that *hsp16* is overexpressed after just 30 minutes of exposure to elevated temperature and persists into 4 hours (Figure Appendix I.1A). I also find that loss of *cid12* without any temperature stress exposure leads to overexpression of *hsp16* that appears as drastic as when wild-type cells were exposed to stress (Figure Appendix I.1A). This suggests that *cid12* plays a dominant role in the silencing of *hsp16* under non-stressed conditions.

To explore whether other factors contribute to *hsp16* silencing, I deleted various genes in *S. pombe* cells and assayed *hsp16* RNA expression under non-stressed conditions. In addition to confirming that loss of *dcr1* leads to *hsp16* overexpression, I found that deletion of the RDRC component Rdp1 also causes overexpression (Figure Appendix I.1B). Moreover, double mutants *cid12Δdcr1Δ* and *cid12Δrdp1Δ* appears to cause *hsp16* overexpression that is not more severe than any of the single mutants, suggesting that Cid12, Dcr1, and Rdp1 function in the same pathway to repress *hsp16* (Figure Appendix I.1B). This result is reminiscent of the role these three proteins play in centromeric silencing (Motamedi et al., 2004). It is likely that, like at the centromere, Cid12 functions in complex with Rdp1 to silence *hsp16*, although more experiments are required to confirm this hypothesis. Additionally, I find that loss of the RITS components, Tas3 or Chp1, also leads to depression of *hsp16* (Figure Appendix I.1C). Surprisingly, I find that loss of the heterochromatin factors Clr4 and Swi6 cause only slight *hsp16* up-regulation (Figure Appendix I.1B). Together, these results suggest that, like at the centromere, RITS complex and



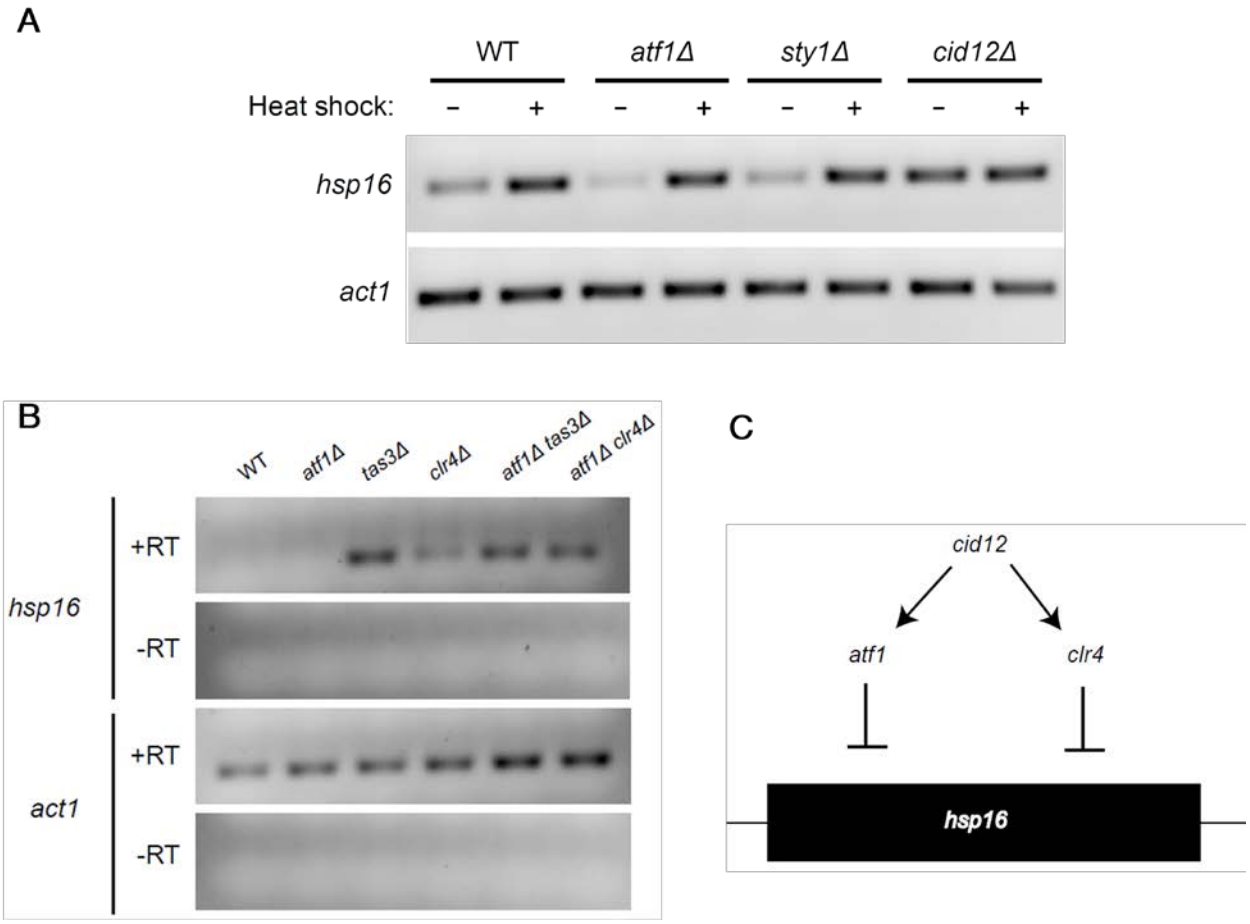
Appendix Figure I.1 Identification of *S. pombe* factors involved in *hsp16* transcriptional repression

(A) Semi-quantitative RT-PCR (Semi-RT-PCR) of wild-type versus *cid12Δ* cells, probing for *hsp16* RNA levels with 0, 30, or 240 minutes of pre-exposure to heatshock at 37°C. (B) Semi-RT-PCR of wild-type or mutant cells, probing for *hsp16* RNA levels in unstressed cells. (C) Semi-RT-PCR of wild-type or mutant cells, probing for *hsp16* or *hsp104* RNA levels in unstressed cells. (D) Temperature-sensitivity cell growth assay of wild-type or mutant cells. Log-phase cells were spread on YES plate and incubated at either 30°C (unstressed) or 37°C (stressed) for 3 days prior to imaging. -RT, no reverse transcriptase. +RT, with reverse transcriptase. *Act1*$^{+}$ serves as loading control. WT denotes wild-type.

RDRC components are involved in *hsp16* silencing. Moreover, unlike at the centromere, heterochromatin is not a dominant factor for silencing here.

In Chapter 3, I discovered that mutations with disrupt the physical interaction between Cid12 and Atf1 caused *hsp16* derepression. Perhaps most surprising, I find that complete loss of *atf1* does not cause *hsp16* overexpression (Figure Appendix I.1B). Counter-intuitively, *atf1Δ* showed even lower levels of *hsp16* RNA as compared to wild-type unstressed cells. Moreover, I find that *atf1Δ* cells appear as viable when grown under heatshock as wild-type cells (Figure Appendix I.1D). It is known that the kinase Sty1 functions upstream of Atf1. After stress exposure, Sty1 transverses from the cytosol into the nucleus where it phosphorylates Atf1 (Wilkinson et al., 1996). Phosphorylated Atf1 is stabilized and can transcriptionally influence downstream target genes (Lawrence et al., 2007). I find that *sty1Δ* cells are hypersensitive to elevated temperatures, suggesting a possible influence of Sty1 on *hsp16* expression (Figure Appendix I.1D). Thus, I also tested the effect of *sty1* deletion on *hsp16* expression. However, I find that *hsp16* RNA levels are unchanged in *sty1Δ* cells (Figure Appendix I.1B).

Traditionally, phosphorylation of Atf1 by Sty1 has been associated with induced downstream gene expression, such as *fbp1*. Therefore, since Atf1 binds to the promoter of *hsp16*, it is not unreasonable to hypothesize that Atf1 is required to drive *hsp16* expression. To directly test the hypothesis that Atf1 and Sty1 are required for *hsp16* expression, I performed RT-PCR in *atf1Δ* and *sty1Δ* cells, with or without heatshock induction. As control, I also assayed *cid12Δ* cells. In contrast to expectation, I find that deletion of either *atf1* or *sty1* has no effect on overexpression of *hsp16* under condition of heatshock (Figure Appendix I.2A). *Hsp16* is depressed in *cid12Δ* cells, with or without heatshock (Figure Appendix I.2A). Thus, Atf1 and Sty1 are dispensable for activation of *hsp16*.



Appendix Figure I.2 *Atf1* works in parallel with *clr4* to repress *hsp16* RNA expression

(A) Semi-RT-PCR of wild-type or mutant cells, probing for *hsp16* RNA levels in unstressed or stressed cells. Stressed cells exposed to 37°C for 30 minutes just prior to cell harvest. (B) Semi-RT-PCR of wild-type or mutant cells, probing for *hsp16* RNA levels in unstressed cells. (C) Proposed model of repressive *hsp16* expression regulation under non-stressed conditions. -RT, no reverse transcriptase. +RT, with reverse transcriptase. *Act1*$^{+}$ serves as loading control. WT denotes wild-type.

Next, I considered whether Atf1 plays a repressive role of *hsp16* in parallel with another factor under non-stress condition. I find that, while individual mutants of *atf1Δ* or *clr4Δ* does not cause major *hsp16* overexpression, the double mutant *atf1Δ clr4Δ* does (Figure Appendix I.2B). As positive controls, I show that *tas3Δ* or *atf1Δtas3Δ* cells overexpress *hsp16* RNA. Thus, the data indicate that Atf1 and Clr4 might be involved in parallel pathways the mediate silencing of *hsp16*.

In Chapter 3, I find that disruption of the Cid12-Atf1 leads to *hsp16*derepression. However, I did not prove that the point mutations only disrupted the Cid12-Atf1 interaction. We predicted the Cid12 point mutations to disrupt the interaction with Atf1 but not with the RDRC components, Rdp1 and Hrr1 (Chapter 3). It is possible that the Cid12 mutants also disrupted other unknown interactions. Indeed, my results suggest that the Cid12 mutants may abrogate the functional role of Cid12 in the Atf1 and Clr4 parallel pathways. I hypothesize that Cid12 functions upstream of both Atf1 and Clr4, explaining the fact that loss of Cid12 leads to constitutive *hsp16* transcriptional up-regulation (Figure Appendix I.2C). Downstream, Atf1 and Clr4 might operate independently and redundantly. Interestingly,Clr4 is required to indirectly stabilize the interaction of RDRC and the RITS complex at the centromere via RITS-binding of heterochromatin (Woolcock et al., 2012). Moreover, at the silent mating-type region in *S. pombe*, Atf1 has been found to interact with Clr4 (Jia et al., 2004). While they are capable of interacting, it is still possible that they are kept apart under certain conditions by other regulatory factors. Further work are required to elucidate the functions of these factors at the *hsp16* locus.

There are several questions that still remain. At the moment, it is unclear the precise roles of other factors such as Rdp1 and Tas3. Most likely, they function at the level of Cid12 since loss of any individual factor leads to *hsp16* overexpression. Moreover, my genetic experiments only

reveal the various factors that are required for *hsp16* silencing under non-stressed conditions and a glimpse into some of their punitive genetic interactions. However, the biochemical and molecular details of the silencing mechanism(s) are unknown. How do the factors associate with each other and to the *hsp16* gene body and/or promoter? Does Clr4 impart silencing through direct or indirect mechanism(s)? Additional biochemical and genetic experiments (ChIP, coIP, etc.) will be required to shed light into these mechanisms of heatshock regulation.

APPENDIX II

A NOVEL METHOD TO BARCODE DNA FOR ORFEOME LIBRARY VALIDATIONS AND INTERACTOME MAPPINGS

SUMMARY

While we have virtually complete “parts lists” for tens of thousands of proteins in each sequenced species, what is missing is a means of experimentally connecting the annotated genome to downstream 'omics' applications. A prerequisite for any such effort is to construct a clone library of all open reading frames (ORFs), often called an ORFeome. To this end, we developed a nucleic acid barcoding approach coupled with next-generation sequencing (PLATE-seq) to enable sensitive, reliable, and cost-effective validations of large ORFeomes. Through a minimalistic barcoding scheme, we can comprehensively and precisely identify all genes and their positions. Moreover, when PLATE-seq is integrated into a yeast two-hybrid (Y2H) interaction screening pipeline, the approach allows rapid identification and validation of interactions. We leverage PLATE-seq to (1) generate a fully-validated, single colony-based DNA ORFeome library corresponding to ~1,800 *Oryza sativa* genes and to (2) generate a high quality protein interactome network comprising proteins encoded by these validated rice genes.

CONTRIBUTIONS

Haiyuan conceived of the initial PLATE-seq idea and supervised research. I performed all experiments and troubleshooting. I generated a single-colony rice entry clone ORFeome by picking 2 single bacteria colonies for each of >3,000 wells. Nurten Akturk and I performed yeast two-hybrid experiments. I performed Sanger sequencing, BLAST analyses, and pairwise Y2H experiments. Peter Schweitzer, Jennifer Mosher, and Ann Tate performed MiSeq runs. Peter Schweitzer was instrumental in assisting with PLATE-seq designs. Robert Fragoza wrote a Python script to generate all 7 bp barcodes which I assembled into primers. Michael Meyer performed all next-generation sequencing analyses.

INTRODUCTION

Large repositories of genes (ORFeomes) serve as essential backbones of all genome-wide functional screens. The ability to directly manipulate genes on a massive scale has enabled comprehensive systematic functional studies including those in search of protein-protein interactions (Das et al., 2013; Vo et al., 2016; Yu et al., 2008). Thus, is it not surprising that huge amounts of effort and resources have been invested in generating these libraries (Bischof et al., 2014; Chai et al., 2015; Grant et al., 2015; Lamesch et al., 2007; Rual et al., 2004; Temple et al., 2009).

ORFeome libraries are commonly generated through gene-specific amplification by polymerase chain-reactions (PCRs) followed by cloning into versatile vectors such as Gateway vectors. While collaborations and the advent of automated high-throughput technologies have drastically facilitated the generation of ORFeome, our ability to fully validate these massive libraries has been lagging. Sanger and next-generation sequencing are the traditional approaches to validate ORFeomes. Sanger sequencing has been the gold standard for ORFeome validation because sequencing from the 5' and 3' ends can generate long and reliable expression sequence tags (ESTs) to sufficiently identify most genes. However, this method suffers from low sensitivity (usually requiring high concentrations of DNA template) and high cost which scales with the number of sequencing reactions. Moreover, this approach cannot identify pools of different DNA species, which would require laborious pre-sorting of all species. Thus, Sanger sequencing becomes increasingly prohibitive as the size and number of ORFeomes increase. In contrast, next-generation sequencing has dramatically reduced the cost of sequencing. This approach permits large-scale pooling of ORFeomes to identify all DNA species with high accuracy and depth. Despite the huge improvement in costs, the generation of pooled DNA precludes the identification of the precise position of every gene within the ORFeome library.

Doing so would require the impractical use of many index adapters. The lost of position information is highly problematic for large-scale studies. Often, genes are cloned or cherrypicked to generate multiple sub-libraries. An inherent issue that commonly arises is that of cross-contamination when dealing with vast amounts of samples. Another possibility arises when multiple pools of genes are handled which require subsequent deconvolution. Thus, a reliable and cost-effective method for comprehensive identification of genes and their exact positions within an ORFeome is a warranted goal.

DNA barcodes have become a promising molecular tool for multiplexed analyses. For instance, it has been used to identify unique yeast mutant strains (Kim et al., 2010) or pathogen DNA (Li et al., 2005). The principle purpose of these barcodes is to identify the source of DNA elements which persists after a screen-of-interest. We leverage this tool to develop PCR-mediated linkage of barcoded adapters to nucleic acid elements for sequencing (PLATE-seq) to generate *de novo* barcodes to precisely and comprehensively validate ORFeomes in high-throughput.

We first apply PLATE-seq to fully-validate an ORFeome consisting of ~3,000 *Oryza sativa* genes. For a library of this size, our method costs 2×10^{-5} times as would be compared to use of the gold standard, Sanger sequencing. Moreover, PLATE-seq generates barcodes *de novo*, which eliminates the requirements for cloning static barcodes within every library and allows the same barcodes to be used across multiple libraries. We also constructed a seamless pipeline that integrates high-throughput yeast two-hybrid (Y2H) with PLATE-seq to generate the first systematic large-scale protein interactome network of *O. sativa*. The resultant network reveals numerous previously unknown interactions which may impact agriculture biology.

MATERIALS AND METHODS

Design of adapter sequences

We required that each plate-specific, position-specific, or clone-specific adapter sequence must differ so that at least two specific point mutations would be required to change one adapter into another. We reasoned that, with an error rate per base of ~10^{-7} for our phusion polymerase enzyme, our threshold of at least two mutations should be sufficient to overcome the possibility of making a false gene-position call due to errors during PCR.

Nextera tagmentation

Illumina Nextera tagmentation reactions were performed according to manufacturer instructions, except I used 1.2uL of commercial Tn5 transposase enzyme per 12.5ng DNA tagmentation. It is important to perform test titrations of the tagmentation reactions and compare Bioanalyzer results to ensure the generation of DNA fragment sizes 300 – 700 bp.

Generation of an *O. sativa* ORFeome

About 3,000 rice genes were initially amplified by PCR using gene-specific primers and cDNA. Then, PCR products were cloned into the Gateway-compatible plasmid pDONR223. Thereafter, clones were transformed into DH5α chemically-competent cells. To ensure that all bacterial constructs contained a single clone of a gene, we spread all bacterial constructs onto selective LB + spectinomycin plates and picked at least two single colonies per well. All single colonies which passed verification by PLATE-seq (*ie*. the clone was of the correct gene and present in the expected plate-position) was arrayed into our final entry clone library.

Yeast two-hybrid

Yeast two-hybrid experiments were performed as detailed in Chapter 3 Materials and Methods. At the conclusion of Y2H Phenotyping II, we picked yeast colonies that were Y2H positive into 96-well plates.

Sanger sequencing and analysis

PCR amplicons were sequenced at the Cornell genomics facility and were aligned using NCBI tBLASTn and the annotated *O. sativa* ORFeome library.

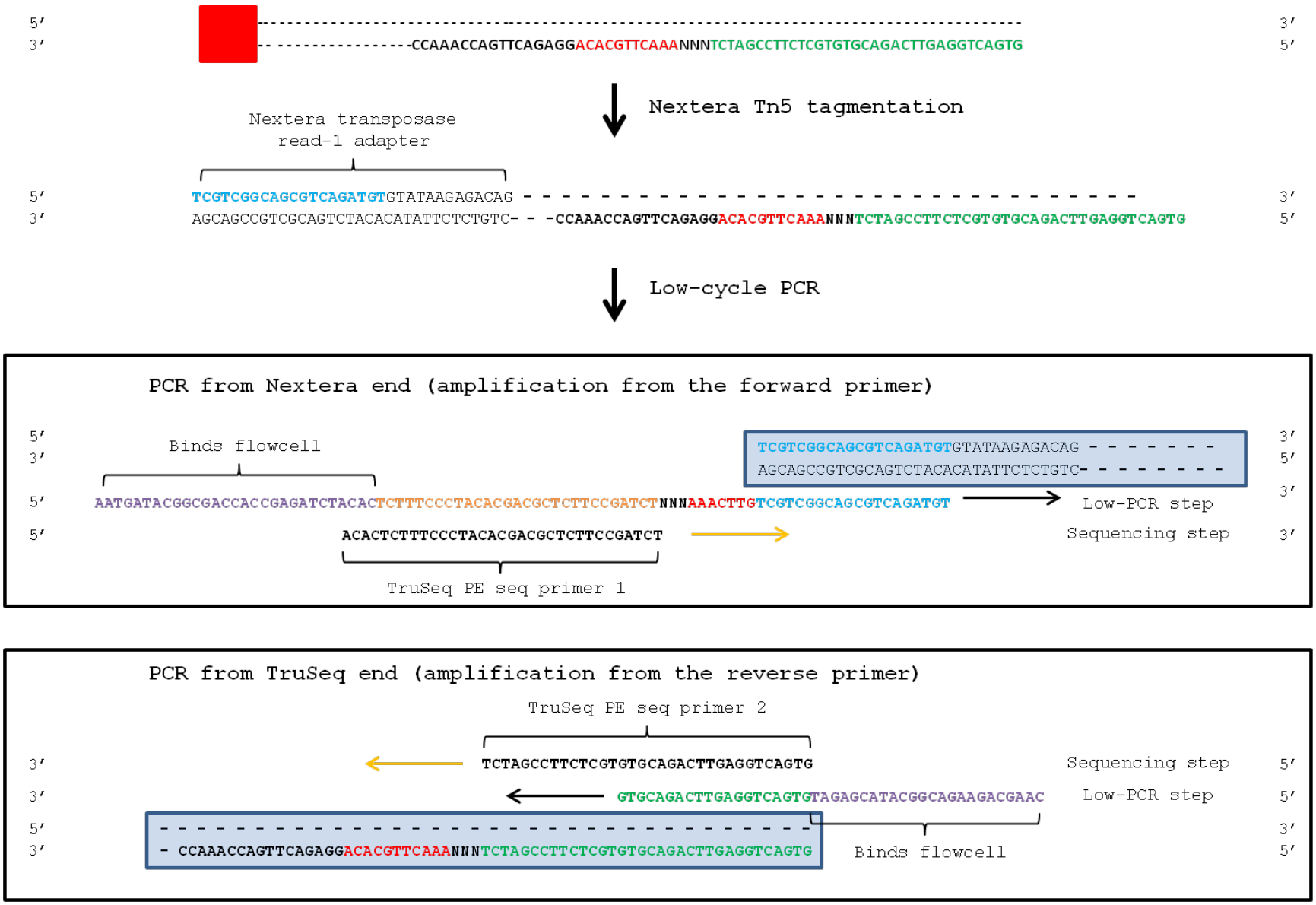
RESULTS AND CONCLUSIONS

The primary premise of PCR-mediated Linkage of Adapter barcodes To nucleic Elements for sequencing (PLATE-seq) is to leverage next-generation sequencing to determine the precise identity and position of any gene-of-interest on a high-throughput scale. To accomplish this, we require at minimum known adapters that enables site-specific sequencing and identifies plate-position information (Appendix II Figure 1).

For a single 96-well plate, where a gene-of-interest is within each well position, we designed a single universal forward primer that recognizes the plasmid backbone just prior to each gene. We designed ninety-six unique reverse primers which consist of a universal reverse priming sequence, a position-specific adapter, and an Illumina TruSeq adapter 2 sequence. Polymerase chain reactions (PCRs) enabled the amplification and barcoding of all genes-of-interest within the plate with position-specific adapters. Position-specific adapters were designed to minimize the identity between any two unique adapters to prevent errors in position

Appendix Figure II.1 Schematic of main PLATE-seq pipeline

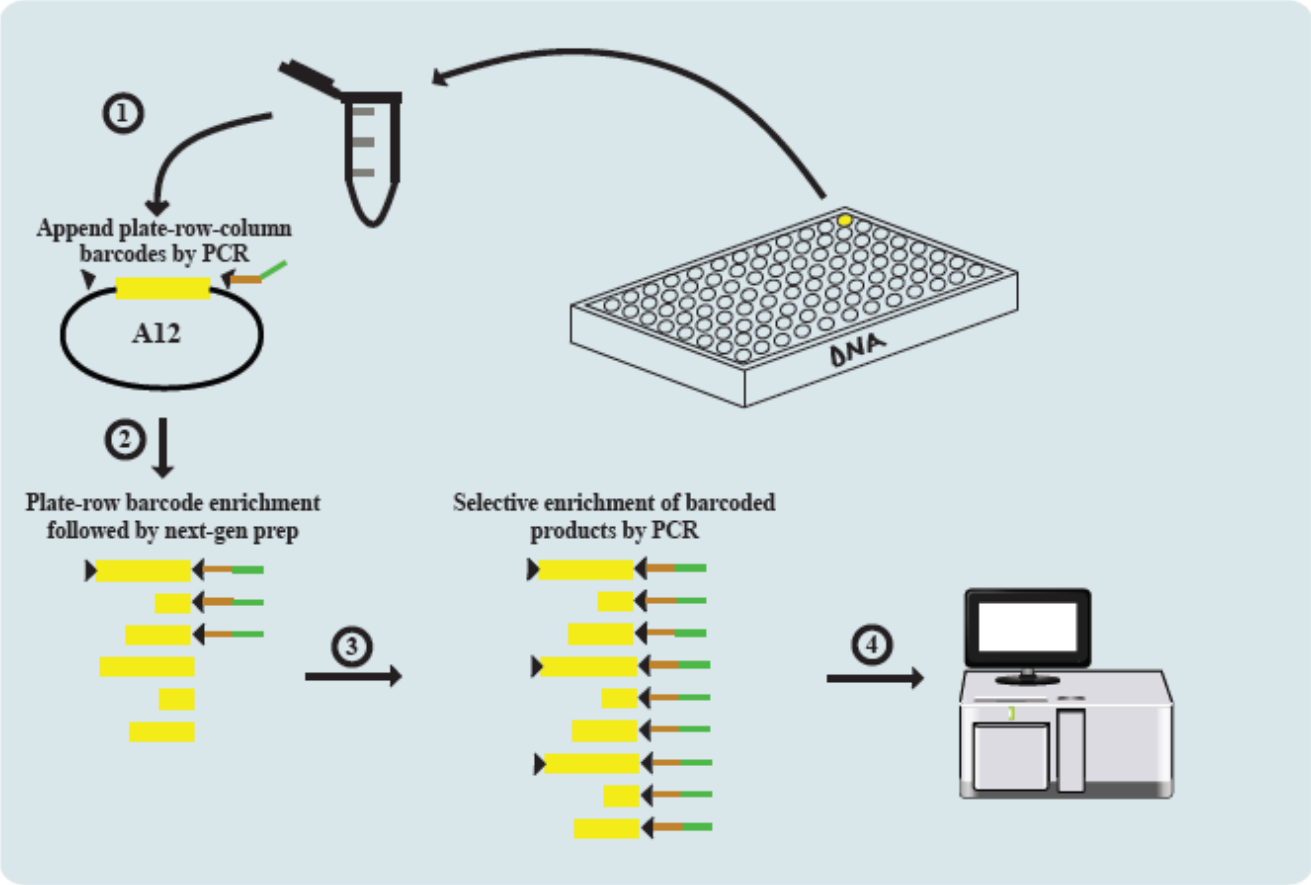
The red box denotes the ORF sequence. Actual sequences in red are example plate or position barcodes. Sequences in orange or green represent the TruSeq adapter 1 or 2, respectively. Sequences in blue are priming sites. Sequences in purple are Illumina flowcell binding sites.


5′
3′
CCAAACCAGTTCAGAGGACACGTTCAAANNNTCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG
Nextera Tn5 tagmentation
Nextera transposase
read-1 adapter
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
AGCAGCCGTCGCAGTCTACACATATTCTCTGTC
Low-cycle PCR
PCR from Nextera end (amplification from the forward primer)
Binds flowcell
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNAAACTTGTCGTCGGCAGCGTCAGATGT
Low-PCR step
ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Sequencing step
TruSeq PE seq primer 1
PCR from TruSeq end (amplification from the reverse primer)
TruSeq PE seq primer 2
TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG
GTGCAGACTTGAGGTCAGTGTAGAGCATACGGCAGAAGACGAAC
CCAAACCAGTTCAGAGGACACGTTCAAANNNTCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG
Binds flowcell

annotations arising from the change of an adapter into another due to mutations introduced during PCR. This process results in a fast and easy way to label the position of every gene.

Next, all barcoded PCR amplicons from a single plate would be pooled together and fragmented through Tn5-mediated tagmentation. The tagmentation procedure allows the ligation of Nextera adapters at the sites of cleavage. Since tagmentation results in random cleavages along the entire body of each amplicon, only a small fraction of amplicon fragments (<5%) are expected to contain part of each gene-of-interest, the corresponding position-specific adapter, and the TruSeq adapter 2 sequence. Through another round of PCR on the tagmented amplicon pool, we extend the Nextera transposase read-1 adapter to include a plate-specific adapter and an Illumina TruSeq adapter 1 sequence. More importantly, during this round of PCR, primers specifically recognize the Nextera transposase read-1 adapter and the TruSeq adapter 2, which leads to preferential amplification of those amplicon fragments consisting of both plate-specific and position-specific adapters along with part of each gene-of-interest and TruSeq sequencing sites. This approach generates DNA fragments containing partial gene sequence along with adapters signifying plate and position identity. Ultimately, we used Illumina next-generation sequencing to identify all genes within a 96-well plate and their precise plate-position.

Our PLATE-seq approach for a single plate can be easily extended for multi-plate DNA libraries. If all plates are of the same dimensions (*eg*. they are all 96-well plates of 12 columns and 8 rows), then only ninety-six unique primers with position-specific adapters are necessary. The number of primers with plate-specific adapters scales linearly with the number of plates. Thus, PLATE-seq is an automatable pipeline that requires the design of few unique primers and can cost-effectively identify all genes within a library and their exact locations.



Appendix Figure II.2 Schematic of alternative PLATE-seq design

The yellow rectangles denote a specific ORF-of-interest. The brown line is sequence that recognizes a site on the plasmid backbone. The green line is sequence that encodes the plate barcode, position barcode, and TruSeq adapter 1 sequence.

Additionally, we tested the possibility of appending all barcodes to the ORF sequences during the initial PCR step (Appendix II Figure 2). To do this, we designed partially complementary primers whereby the plate barcode would be on one primer and the position barcode would be on the other. During an initial PCR step, the complementary primers would anneal and extend each other to generate dsDNA consisting of plate and position barcodes. This product would serve a primer for a secondary round of PCR targeting each ORF sequence. At the conclusion of both PCR steps, an amplicon consisting of the full-length ORF, plate and position barcodes, and a TruSeq adapter 2 sequence would be obtained. These amplicons were subject to Nextera prep by tagmentation and sequenced on a MiSeq platform.

To test this alternative approach, I began with a 96-well plate where each well consisted of a unique purified plasmid (consisting different human ORFs in vectors pDEST AD, pDEST DB, or pDONR223) (Appendix II Table 2). All plasmids were Sanger sequenced. Using PLATE-seq, we were able to identify ~95% of all test ORFs with correctly associated plate positions. For every position, the number of reads corresponding with the correct ORF (in pDEST AD or DB) was 100 to 1000 times greater than the number of reads calling another incorrect ORF (compared with the incorrect ORF with the greatest number of supporting reads). The signal-to-noise separation was much lower for calling ORFs in pDONR223 (10:1 to 50:1), most likely because of poor PCR efficiencies. However, I found that this strategy was incompatible with plasmids from yeast lysates, which are not purified (Appendix II Figure 3). I found that the purity of the plasmids greatly affected the efficiency of the PCRs required to link ORFs and barcodes. To prove that the nature of my dsDNA (containing the plate and position barcodes) was the source of problems in PCRs, I synthesized ssDNA with the same barcodes. Using plasmids from yeast lysates, I found that PCR using the ssDNA as primer was much more efficient than using dsDNA

ORF	PlateNo	Position
14525	1	A01
7444	1	B01
7443	1	C01
3552	1	D01
14580	1	E01
56166	1	F01
292	1	G01
817	1	H01
1686	1	A02
10558	1	B02
5073	1	C02
6080	1	D02
3910	1	E02
2963	1	F02
6557	1	G02
7150	1	H02
3887	1	A03
3117	1	B03
55759	1	C03
3961	1	D03
3128	1	E03
4185	1	F03
4137	1	G03
5006	1	H03
4140	1	A04
4665	1	B04
9000	1	C04
7034	1	D04
3261	1	E04
1362	1	F04
8297	1	G04
7060	1	H04
42	1	A05
4707	1	B05
5591	1	C05
2409	1	D05
2555	1	E05
6754	1	F05
5675	1	G05

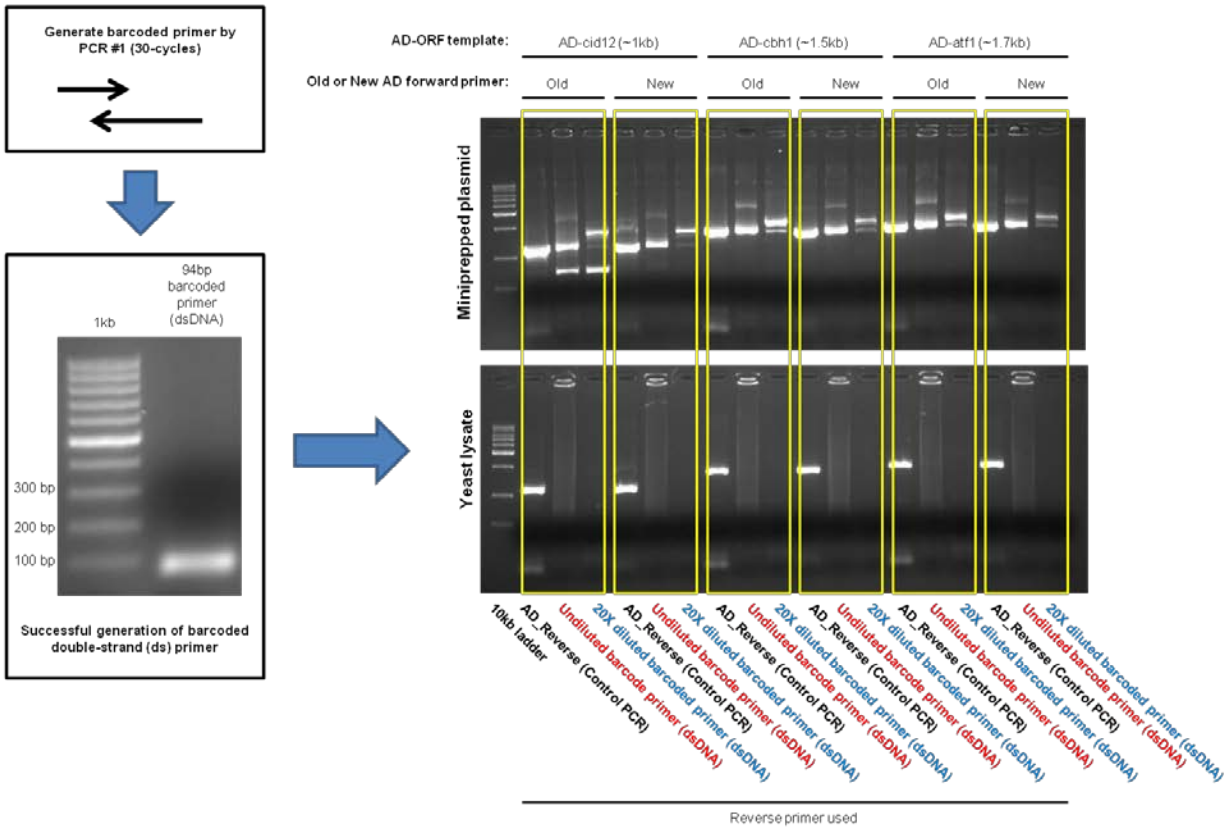
Appendix Table II.1 List of test-case human ORFs in 96-well plate format

8464	1	H05
8010	1	A06
7982	1	B06
7886	1	C06
5513	1	D06
5456	1	E06
5498	1	F06
3801	1	G06
10033	1	H06
3182	1	A07
4019	1	B07
8905	1	C07
2709	1	D07
280	1	E07
1256	1	F07
783	1	G07
311	1	H07
7022	1	A08
836	1	B08
8296	1	C08
53836	1	D08
2324	1	E08
322	1	F08
3971	1	G08
7127	1	H08
365	1	A09
6300	1	B09
1364	1	C09
3154	1	D09
4044	1	E09
54902	1	F09
11805	1	G09
54914	1	H09
596	1	A10
5244	1	B10
53021	1	C10
7631	1	D10
7606	1	E10
7762	1	F10
6699	1	G10

(Appendix Table II.2 continued)

7109	1	H10
14679	1	A11
11943	1	B11
10017	1	C11
11706	1	D11
11372	1	E11
8508	1	F11
4780	1	G11
14758	1	H11
11713	1	A12
5243	1	B12
2278	1	C12
4741	1	D12
3425	1	E12
4687	1	F12
3484	1	G12
3431	1	H12

(Appendix Table II.2 continued)



Appendix Figure II.3 Long (>90bp) dsDNA primers do not work on PCRs of plasmid templates in yeast lysate but work well using purified plasmids

As test cases, I used three different pDEST AD plasmids containing *S. pombe* ORFs *cid12*, *cbh1*, or *atf1*. First, I generated dsDNA with plate and position barcodes through a first round of PCR. I show, by agarose gel analysis, that the first PCR yield product (expected size of 94bp). Next, I used the dsDNA as primer for a second round of PCR where the template was either purified plasmid or plasmid from yeast lysate (unpurified). For the second round of PCR, my positive control was PCR using ssDNA primer without any barcodes and ~25bp in size. Aside from the control PCR, plasmids in yeast lysate yield no PCR product while PCR products appear when using purified plasmid in PCR with dsDNA primers. I also tried 20X dilutions of the dsDNA prior to setting up the second round PCRs.

as primer, although not as efficient as using purified plasmids. In trying to circumvent the ssDNA versus dsDNA issue, I tried to generate dsDNA with 5' phosphorylation labeling the unnecessary strand for downstream digestion. However, although I could turn dsDNA into ssDNA, this did not sufficiently increase my PCR efficiencies. I concluded that I could not integrate this approach with my Y2H pipeline for identifying ORF pairs encoding interacting proteins (because such an approach would need to be able to identify ORFs in plasmids from yeast lysates).

In conclusion, PLATE-seq is still a work-in-progress. I am still trying to optimize barcoding and PCR conditions. While my various approaches were quite successful using high-purity plasmids, they mostly failed when I used yeast lysates containing plasmids. The most likely explanations are that (1) plasmid concentration is very low in the yeast cells with ~1 copy per cell due to the presence of *ARS/CEN* and (2) yeast lysates are also muddled with yeast proteins and genomic DNA. To add to the complexity, the primers required for PLATE-seq are longer than the conventional 20-25bp normally used for Sanger sequencing or normal PCRs, although we require far fewer primers in-total. Design of these primers require careful design to probably require us to take into account possible secondary structures that might inhibit PCRs. To-date, I find that my first PLATE-seq design (which links TruSeq adapters 1 and 2 sequences to both sides of every ORF) appears most promising because, through agarose gel analyses, I find that most PCRs are successful using yeast lysates as templates. However, at the moment, I am having issues with either linking the TruSeq adapter 1 and plate barcode after the tagmentation step or performing next-generation sequencing using the TruSeq adapter 1 as a sequencing site. I am in the process of troubleshooting those details and optimizing the protocols/designs. The overall goal is, once I have an optimized PLATE-seq protocol and design,

I will use it to sequence every entry clone from my rice ORFeome and to integrate into my Y2H pipeline for identifying interacting proteins

APPENDIX III

GENERATION OF A HIGH-DENSITY FUNCTIONAL PROTEIN MICROARRAY

SUMMARY

To-date, yeast two-hybrid is, arguable, the most practical assay for ultra high-throughput screening for binary protein-protein interactions. However, this assay is ~20 years old and has remained relatively unchanged (Fields and Song, 1989). Thus, the technological aspect of protein interactome studies remains lagging. In this Appendix, I discuss roughly 2-years worth of attempts at fabricating a high-density protein microarray.

CONTRIBUTIONS

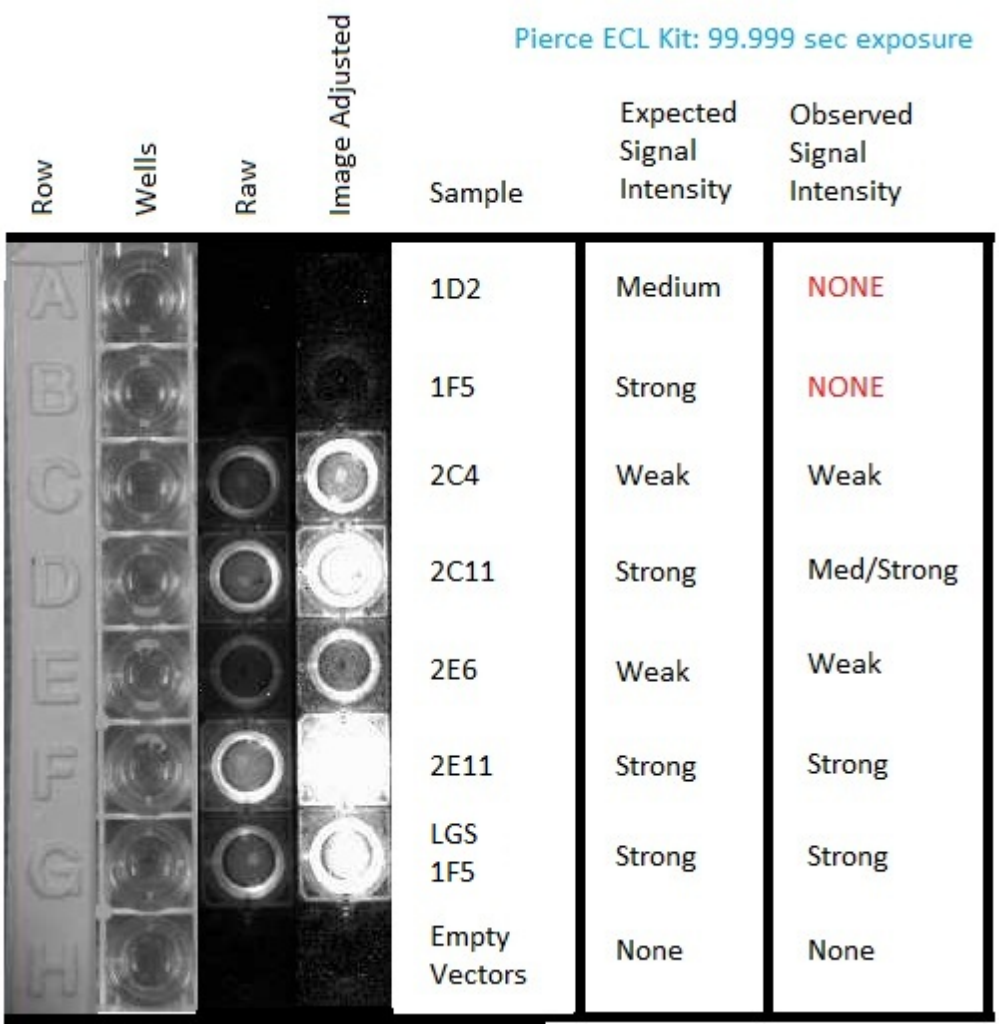
Haiyuan Yu, David Lin, and I were involved in designing and interpreting experiments. Adam Bisogni and I performed experiments.

RESULTS AND CONCLUSIONS

We initially based our protein microarray design from protocols laid forth by Joshua LaBaer and colleagues (Link and Labaer, 2008a, b, c, d, e, f, g, h, i, j; Ramachandran et al., 2004). For all experiments, I used test plasmids of pANT7-nHA or pANT7-cGST with human ORFs. These plasmids derived from Haiyuan's human nucleic acid programmable protein array (NAPPA) positive-control library. I had tested pairs of these plasmids using the well-based NAPPA assay to show which pairs were positive or negative for interaction (Appendix III Figure 1). I made sure to have at least 2 positive and 2 negative controls for the test experiments of the protein microarray designs.

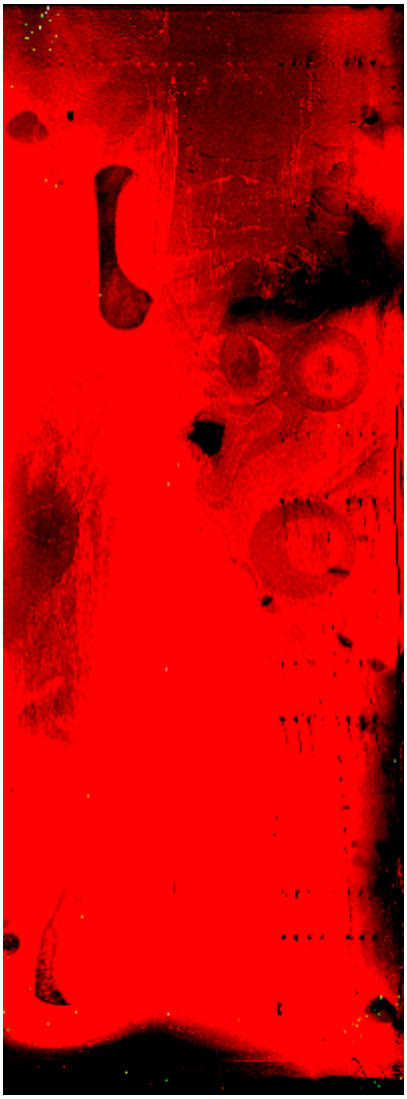
First, I tested to make sure we could print and stick plasmid DNA onto aminosilane or glass slides, effectively to produce DNA microarrays. We printed the plasmids in various buffers containing either 40% glycerol, 1-100mg/mL bovine serum albumin (BSA), water, Pierce Superblock or sarkosyl. After printing, slides were allowed or not allowed to dry overnight with or without humidification. Sybr green staining and imaging of the slides at 532nm showed that plasmid DNA can stick. Moreover, we could wash the slide at least 3-5 times with 1X PBST without washing away most of the DNA.

Next, we printed purified glutathione S-transferase (GST) protein or antibodies onto slides to test protein detection. We tried several methods: either use biotin-avidin to bind proteins to aminosilane-coated glass slides or sticking proteins to epoxy-coated glass slides. With the epoxy-coated slides, we had to optimize blocking and washing conditions to prevent non-specific probe antibody bindings to the slide. In most cases, I found that antibody probing of bound protein was good especially when I hand-printed the spots onto the slide. However, I had much noisier results with machine, micro-scale spots that were much harder to reproduce. I had a lot



Appendix Figure III.1 Well-based NAPPA assay of control pairs of GST or HA-tagged human proteins

I show that samples named 2C4, 2C11, 2E6, 2E11, and LGS 1F5 are NAPPA positives, of variable intensities. Negative controls 1D2, 1F5, and empty vectors gave no signal. These sample pairs would serve as my test cases in attempts to test various designs of the protein microarray. The theory was that the chemistry of the protein microarray should be very similar to that in the well-based NAPPA and therefore, the results should be similar.



Appendix Figure III.2 Detection of GST-tagged protein synthesized *de novo* by TnT reaction and flowed onto glass slide

of trouble getting HA-targeting antibodies to work reliably but did not get new ones due to cost and my uncertainty with antibody qualities. I decided to stick to the antibodies described in (Link and Labaer, 2008a, b, c, d, e, f, g, h, i, j; Ramachandran et al., 2004) until I could exhaust all other factors problematic.

Finally, we tried to synthesize proteins *de novo* on the slides from printed plasmid DNA. We co-spotted capture GST antibodies and plasmids. Then, we tried different ways to add Promega's Transcription/Translation (TnT) commercial mixture of rabbit reticuolysates. The mixture dries rapidly so all incubations had to be done in closed chamber or humidified chamber. Probing for proteins on slide were detected using anti-HA antibodies and secondary with Cy3 fluor. We observed, again, noisy results. Sometimes there were high background. Often, spots gave no signals, indicating lack of protein production and/or protein recognition. To make sure that the TnT mixture worked to produce protein, I performed the protein *de novo* synthesis of GST-tagged proteins in tubes according to manufacturer recommendations. Adam Bisogni confirmed protein expression by GST western blotting and Comassie staining. Furthermore, after protein production in-tube, I added the mixtures onto glass slides and probed for the proteins. I repeatedly observe Cy5 signal on the entire slide, also indicating that there is protein that I can detect on slide (Appendix III Figure 2). Other times, I do see signal, but I see signal in all spots including my negative control spots.

In conclusion, we were unable to fabricate a functional protein microarray. Even after following published protocols many times, we could not reproduce their results. We also tried numerous design variations over the course of 2 years with minimal success. In hindsight, I believe antibody efficiency should be revisited again. In the future, we should more carefully test

the efficiency of multiple different antibodies. It might have been the case that we were stuck with a bad batch of antibodies which hindered all downstream applications.

REFERENCES

Arabidopsis Interactome Mapping Consortium (2011). Evidence for network evolution in an Arabidopsis interactome map. Science *333*, 601-607.

Arellano, M., Coll, P.M., Yang, W., Duran, A., Tamanoi, F., and Perez, P. (1998). Characterization of the geranylgeranyl transferase type I from Schizosaccharomyces pombe. Mol Microbiol *29*, 1357-1367.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Bader, G.D., and Hogue, C.W.V. (2002). Analyzing yeast protein–protein interaction data obtained from different sources. *Nat Biotechnol 20*, 991-997.

Bader, J.S., Chaudhuri, A., Rothberg, J.M., and Chant, J. (2004). Gaining confidence in high-throughput protein interaction networks. Nat Biotechnol *22*, 78-85.

Bajic, D., Moreno-Fenoll, C., and Poyatos, J.F. (2014). Rewiring of genetic networks in response to modification of genetic background. Genome Biol Evol *6*, 3267-3280.

Balasubramanian, M.K., Bi, E., and Glotzer, M. (2004). Comparative Analysis of Cytokinesis in Budding Yeast, Fission Yeast and Animal Cells. Curr Biol *14*, R806-R818.

Barabasi, A.-L., and Oltvai, Z.N. (2004). Network Biology: Understanding the cell's functional organization. *Nat Rev Genet 5*, 101-113.

Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. Science *286*, 509-512.

Ben-Hur, A., and Noble, W.S. (2006). Choosing negative examples for the prediction of protein-protein interactions. BMC Bioinformatics *7 Suppl 1*, S2.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res *28*, 235-242.

Bischof, J., Sheils, E.M., Bjorklund, M., and Basler, K. (2014). Generation of a transgenic ORFeome library in Drosophila. Nat Protoc *9*, 1607-1620.

Blencowe, B.J., Issner, R., Nickerson, J.A., and Sharp, P.A. (1998). A coactivator of pre-mRNA splicing. Genes Dev *12*, 996-1009.

Bone, N., Millar, J.B., Toda, T., and Armstrong, J. (1998). Regulated vacuole fusion and fission in Schizosaccharomyces pombe: an osmotic response dependent on MAP kinases. Curr Biol *8*, 135-144.

Braun, P., Tasan, M., Dreze, M., Barios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J.M., Murray, R.R., Roncari, L., Smet, A.-S.d., *et al.* (2008). An experimentally derived confidence score for binary protein-protein interactions. Nat Methods *6*, 91-97.

Breiman, L. (2001). Random forests. Mach Learn *45*, 5-32.

Causier, B. (2004). Studying the interactome with the yeast two-hybrid system and mass spectrometry. Mass Spectrom Rev *23*, 350-367.

Ceol, A., Chatr Aryamontri, A., Licata, L., Peluso, D., Briganti, L., Perfetto, L., Castagnoli, L., and Cesareni, G. (2010). MINT, the molecular interaction database: 2009 update. Nucleic Acids Res *38*, D532-539.

Chai, C., Wang, Y., Joshi, T., Valliyodan, B., Prince, S., Michel, L., Xu, D., and Nguyen, H.T. (2015). Soybean transcription factor ORFeome associated with drought resistance: a valuable resource to accelerate research on abiotic stress resistance. BMC Genomics *16*, 596.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., *et al.* (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res *40*, D700-705.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404-416.

Conant, G.C., and Wolfe, K.H. (2006). Functional partitioning of yeast co-expression networks after genome duplication. PLoS Biol *4*, e109.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., *et al.* (2010). The genetic landscape of a cell. Science *327*, 425-431.

Cusick, M.E., Yu, H., Smolyar, A., Venkatesan, K., Carvunis, A.-R., Simonis, N., Rual, J.-F., Borick, H., Braun, P., Dreze, M., *et al.* (2009). Literature-curated protein interaction datasets. Nat Methods *6*, 39-46.

Das, J., Mohammed, J., and Yu, H. (2012). Genome-scale analysis of interaction dynamics reveals organization of biological networks. Bioinformatics *28*, 1873-1878.

Das, J., Vo, T.V., Wei, X., Mellor, J.C., Tong, V., Degatano, A.G., Wang, X., Wang, L., Cordero, N.A., Kruer-Zerhusen, N., *et al.* (2013). Cross-species protein interactome mapping reveals species-specific wiring of stress response pathways. Sci Signal *6*, ra38.

Das, J., and Yu, H. (2012). HINT: High-quality protein interactomes and their applications in understanding human disease. BMC Syst Biol *6*, 92.

Dennis, M.Y., Nuttle, X., Sudmant, P.H., Antonacci, F., Graves, T.A., Nefedov, M., Rosenfeld, J.A., Sajjadian, S., Malig, M., Kotkiewicz, H., *et al.* (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. Cell *149*, 912-922.

Eshaghi, M., Lee, J.H., Zhu, L., Poon, S.Y., Li, J., Cho, K.H., Chu, Z., Karuturi, R.K., and Liu, J. (2010). Genomic binding profiling of the fission yeast stress-activated MAPK Sty1 and the bZIP transcriptional activator Atf1 in response to H2O2. PLoS One *5*, e11620.

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

Fares, M.A., Keane, O.M., Toft, C., Carretero-Paulet, L., and Jones, G.W. (2013). The roles of whole-genome and small-scale duplications in the functional specialization of Saccharomyces cerevisiae genes. PLoS Genet *9*, e1003176.

Ficarro, S.B., McCleland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., and White, F.M. (2002). Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotechnol *20*, 301-305.

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

Finn, R.D., Marshall, M., and Bateman, A. (2005). iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. Bioinformatics *21*, 410-412.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., *et al.* (2010). The Pfam protein families database. Nucleic Acids Res *38*, D211-222.

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res *43*, D805-811.

Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., *et al.* (2005). Protein interaction mapping: a Drosophila case study. Genome Res *15*, 376-384.

Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., and Feldman, M.W. (2002). Evolutionary rate in the protein interaction network. Science *296*, 750-752.

Frost, A., Elgort, M.G., Brandman, O., Ives, C., Collins, S.R., Miller-Vedam, L., Weibezahn, J., Hein, M.Y., Poser, I., Mann, M., *et al.* (2012). Functional repurposing revealed by comparing S. pombe and S. cerevisiae genetic interactions. Cell *149*, 1339-1352.

Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., *et al.* (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216-220.

Gandhi, T.K., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K.N., Mohan, S.S., Sharma, S., Pinkert, S., Nagaraju, S., Periaswamy, B., *et al.* (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet *38*, 285-293.

Gasch, A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. Yeast *24*, 961-976.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., *et al.* (2001). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature 415*, 141-147.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., *et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.

Gibson, T.A., and Goldberg, D.S. (2009). Questioning the ubiquity of neofunctionalization. PLoS Comput Biol *5*, e1000252.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., *et al.* (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.

Goh, C.S., Bogan, A.A., Joachimiak, M., Walther, D., and Cohen, F.E. (2000). Co-evolution of proteins with their interaction partners. J Mol Biol *299*, 283-293.

Grant, I.M., Balcha, D., Hao, T., Shen, Y., Trivedi, P., Patrushev, I., Fortriede, J.D., Karpinka, J.B., Liu, L., Zorn, A.M., *et al.* (2015). The Xenopus ORFeome: A resource that enables functional genomics. Dev Biol *408*, 345-357.

Guan, Y., Dunham, M.J., and Troyanskaya, O.G. (2007). Functional analysis of gene duplications in Saccharomyces cerevisiae. Genetics *175*, 933-943.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Hakes, L., Lovell, S.C., Oliver, S.G., and Robertson, D.L. (2007a). Specificity in protein interactions and its relationship with sequence diversity and coevolution. Proc Natl Acad Sci U S A *104*, 7999-8004.

Hakes, L., Pinney, J.W., Lovell, S.C., Oliver, S.G., and Robertson, D.L. (2007b). All duplicates are not equal: the difference between small-scale and genome duplication. Genome Biol *8*, R209.

He, X., and Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics *169*, 1157-1164.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Luhrmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell *45*, 567-580.

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A *89*, 10915-10919.

Hirsh, A.E., and Fraser, H.B. (2001). Protein dispensability and rate of evolution. Nature *411*, 1046-1049.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., *et al.* (2001). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature 415*, 180-183.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet *16*, 71-84.

Hu, Z., Ng, D.M., Yamada, T., Chen, C., Kawashima, S., Mellor, J., Linghu, B., Kanehisa, M., Stuart, J.M., and DeLisi, C. (2007). VisANT 3.0: new modules for pathway visualization, editing, prediction and construction. Nucleic Acids Res *35*, W625-632.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686-691.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., *et al.* (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell *162*, 425-440.

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

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., and Gerstein, M. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science *302*, 449-453.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. Nature *407*, 651-654.

Jia, S., Noma, K., and Grewal, S.I. (2004). RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. Science *304*, 1971-1976.

Johnson, A.E., Chen, J.S., and Gould, K.L. (2013). CK1 is required for a mitotic checkpoint that delays cytokinesis. Curr Biol *23*, 1920-1926.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., and Burley, S.K. (2001). Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. Cell *106*, 71-81.

Kasahara, M. (2007). The 2R hypothesis: an update. Curr Opin Immunol *19*, 547-552.

Kastritis, P.L., Moal, I.H., Hwang, H., Weng, Z., Bates, P.A., Bonvin, A.M., and Janin, J. (2011). A structure-based benchmark for protein-protein binding affinity. Protein Sci *20*, 482-491.

Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature *428*, 617-624.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., *et al.* (2012). The IntAct molecular interaction database in 2012. Nucleic Acids Res *40*, D841-846.

Kim, D.U., Hayles, J., Kim, D., Wood, V., Park, H.O., Won, M., Yoo, H.S., Duhig, T., Nam, M., Palmer, G., *et al.* (2010). Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol *28*, 617-623.

Kim, P.M., Lu, L.J., Xia, Y., and Gerstein, M.B. (2006). Relating three-dimensional structures to protein networks provides evolutionary insights. Science *314*, 1938-1941.

Kim, W.K., Bolser, D.M., and Park, J.H. (2004). Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). Bioinformatics *20*, 1138-1150.

Kleinberg, E. (1996). An overtraining-resistant stochastic modeling method for pattern recognition. Ann Stat *24*, 2319–2349.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., *et al.* (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.

Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P., *et al.* (2007). hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics *89*, 307-315.

Lawrence, C.L., Maekawa, H., Worthington, J.L., Reiter, W., Wilkinson, C.R., and Jones, N. (2007). Regulation of Schizosaccharomyces pombe Atf1 protein levels by Sty1-mediated phosphorylation and heterodimerization with Pcr1. J Biol Chem *282*, 5160-5170.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.-O., Han, J.-D.J., Chesneau, A., Hao, T., *et al.* (2004). A Map of the Interactome Network of the Metazoan C. elegans. Science *303*, 540-543.

Li, Y., Cu, Y.T., and Luo, D. (2005). Multiplexed detection of pathogen DNA with DNA-based fluorescence nanobarcodes. Nat Biotechnol *23*, 885-889.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Link, A.J., and Labaer, J. (2008a). Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 1: Coating Glass Slides with Amino Silane. CSH Protoc *2008*, pdb prot5056.

Link, A.J., and Labaer, J. (2008b). Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 2: Preparing Bacterial Cultures in a 96-Well Format. CSH Protoc *2008*, pdb prot5057.

Link, A.J., and Labaer, J. (2008c). Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 3: Isolating DNA Plasmids in a 96-Well Plate Format. CSH Protoc *2008*, pdb prot5058.

Link, A.J., and Labaer, J. (2008d). Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 4: DNA Biotinylation, Precipitation, and Arraying of Samples. CSH Protoc *2008*, pdb prot5059.

Link, A.J., and Labaer, J. (2008e). Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 5: Expressing Proteins on NAPPA Slides. CSH Protoc *2008*, pdb prot5060.

Link, A.J., and Labaer, J. (2008f). Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 6: Detecting Proteins on NAPPA Slides. CSH Protoc *2008*, pdb prot5061.

Link, A.J., and Labaer, J. (2008g). Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 7: Detecting DNA on NAPPA Slides. CSH Protoc *2008*, pdb prot5062.

Link, A.J., and Labaer, J. (2008h). Using the Nucleic Acid Programmable Protein Array (NAPPA) for Identifying Protein-Protein Interactions. Protocol 1: Coexpression of Query Protein on NAPPA Slides. CSH Protoc *2008*, pdb prot5108.

Link, A.J., and Labaer, J. (2008i). Using the Nucleic Acid Programmable Protein Array (NAPPA) for Identifying Protein-Protein Interactions. Protocol 2: Detection of Query Proteins on NAPPA Slides. CSH Protoc *2008*, pdb prot5109.

Link, A.J., and Labaer, J. (2008j). Using the Nucleic Acid Programmable Protein Array (NAPPA) for Identifying Protein-Protein Interactions: General Guidelines. CSH Protoc *2008*, pdb ip62.

Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science *286*, 295-299.

Matsuyama, A., Arai, R., Yashiroda, Y., Shirai, A., Kamata, A., Sekido, S., Kobayashi, Y., Hashimoto, A., Hamamoto, M., Hiraoka, Y., *et al.* (2006). ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol *24*, 841-847.

Matthews, L.R., Vaglio, P., Reboul, J., Ge, H., Davis, B.P., Garrels, J., Vincent, S., and Vidal, M. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". Genome Res *11*, 2120-2126.

McDowall, M.D., Harris, M.A., Lock, A., Rutherford, K., Staines, D.M., Bahler, J., Kersey, P.J., Oliver, S.G., and Wood, V. (2015). PomBase 2015: updates to the fission yeast database. Nucleic Acids Res *43*, D656-661.

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

Mewes, H.W., Albermann, K., Bähr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., K. Kleine, Maierl, A., Oliver, S.G., *et al.* (1997). Overview of the yeast genome. Nature *387*, 7-8.

Mewes, H.W., Ruepp, A., Theis, F., Rattei, T., Walter, M., Frishman, D., Suhre, K., Spannagl, M., Mayer, K.F., Stumpflen, V., *et al.* (2011). MIPS: curated databases and comprehensive secondary data resources in 2010. Nucleic Acids Res *39*, D220-224.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol *48*, 443-453.

Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol *3*, 418-426.

Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods *9*, 471-472.

Oliver, S. (2000). Guilt-by-association goes global. *Nature 403*, 601-603.

Papp, B., Pal, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. Nature *424*, 194-197.

Petronczki, M., Matos, J., Mori, S., Gregan, J., Bogdanova, A., Schwickart, M., Mechtler, K., Shirahige, K., Zachariae, W., and Nasmyth, K. (2006). Monopolar attachment of sister kinetochores at meiosis I requires casein kinase 1. Cell *126*, 1049-1064.

Pidoux, A.L., Richardson, W., and Allshire, R.C. (2003). Sim4: a novel fission yeast kinetochore protein required for centromeric silencing and chromosome segregation. J Cell Biol *161*, 295-307.

Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res *37*, 825-831.

Qian, J., Dolled-Filhart, M., Lin, J., Yu, H., and Gerstein, M. (2001). Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. J Mol Biol *314*, 1053-1066.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A.Y., Walter, J.C., and LaBaer, J. (2004). Self-assembling protein microarrays. Science *305*, 86-90.

Remy, I., and Michnick, S.W. (2006). A highly sensitive protein-protein interaction assay based on Gaussia luciferase. Nat Methods *3*, 977-979.

Rensing, S.A. (2014). Gene duplication as a driver of plant morphogenetic evolution. Curr Opin Plant Biol *17*, 43-48.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

Rhind, N., Chen, Z., Yassour, M., Thompson, D.A., Haas, B.J., Habib, N., Wapinski, I., Roy, S., Lin, M.F., Heiman, D.I., *et al.* (2011). Comparative functional genomics of the fission yeasts. Science *332*, 930-936.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet *16*, 276-277.

Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., and Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal *4*, rs3.

Riley, R., Lee, C., Sabatti, C., and Eisenberg, D. (2005). Inferring protein domain interactions from databases of interacting proteins. Genome Biol *6*, R89.

Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.-O., Hayles, J., *et al.* (2008). Conservation and rewiring of functional modules revealed by an epistatis map in fission yeast. Science *322*, 405-410.

Rolland, T., Tasan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., *et al.* (2014). A proteome-scale map of the human interactome network. Cell *159*, 1212-1226.

Rual, J.F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.O., *et al.* (2004). Human ORFeome version 1.1: a platform for reverse proteomics. Genome Res *14*, 2128-2135.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., *et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.

Rustici, G., Mata, J., Kivinen, K., Lio, P., Penkett, C.J., Burns, G., Hayles, J., Brazma, A., Nurse, P., and Bahler, J. (2004). Periodic gene expression program of the fission yeast cell cycle. Nat Genet *36*, 809-817.

Ryan, C.J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S.R., *et al.* (2012). Hierarchical modularity and the evolution of genetic interactomes across species. Mol Cell *46*, 691-704.

Sahni, N., Yi, S., Taipale, M., Fuxman Bass, J.I., Coulombe-Huntington, J., Yang, F., Peng, J., Weile, J., Karras, G.I., Wang, Y., *et al.* (2015). Widespread macromolecular interaction perturbations in human genetic disorders. Cell *161*, 647-660.

Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. Nucleic Acids Res *32*, D449-451.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowsk, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res 13*, 2498-2504.

Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., McCuine, S., Uetz, P., Sittler, T., Karp, R.M., and Ideker, T. (2004). Conserved patterns of protein interaction in multiple species. *PNAS 102*, 1974-1979.

Shevchenko, A., Roguev, A., Schaft, D., Buchanan, L., Habermann, B., Sakalar, C., Thomas, H., Krogan, N.J., and Stewart, A.F. (2008). Chromatin Central: towards the comparative proteome by accurate mapping of the yeast proteomic environment. Genome Biol *9*, R167.

Shiozaki, K., and Russell, P. (1996). Conjugation, meiosis, and the osmotic stress response are regulated by Spc1 kinase through Atf1 transcription factor in fission yeast. Genes Dev *10*, 2276-2288.

Shou, C., Bhardwaj, N., Lam, H.Y., Yan, K.K., Kim, P.M., Snyder, M., and Gerstein, M.B. (2011). Measuring the evolutionary rewiring of biological networks. PLoS Comput Biol *7*, e1001050.

Simonis, N., Rual, J.F., Carvunis, A.R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Sahalie, J.M., Venkatesan, K., Gebreab, F., *et al.* (2009). Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. Nat Methods *6*, 47-54.

Singh, R., Xu, J., and Berger, B. (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. Proc Natl Acad Sci U S A *105*, 12763-12768.

Sipiczki, M. (2000). Where does fission yeast sit on the tree of life? Genome Biol *1*, REVIEWS1011.

Sonnhammer, E.L., and Ostlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res *43*, D234-239.

Stark, C., Breitkreutz, B.J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M.S., Nixon, J., Van Auken, K., Wang, X., Shi, X., *et al.* (2011). The BioGRID Interaction Database: 2011 update. Nucleic Acids Res *39*, D698-704.

Stein, A., Ceol, A., and Aloy, P. (2011). 3did: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res *39*, D718-723.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., *et al.* (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell *122*, 957-968.

Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., and Cooper, D.N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet *133*, 1-9.

Talavera, D., Lovell, S.C., and Whelan, S. (2015). Covariation Is a Poor Measure of Molecular Coevolution. Mol Biol Evol *32*, 2456-2468.

Tardiff, D.F., Jui, N.T., Khurana, V., Tambe, M.A., Thompson, M.L., Chung, C.Y., Kamadurai, H.B., Kim, H.T., Lancaster, A.K., Caldwell, K.A., *et al.* (2013). Yeast reveal a "druggable" Rsp5/Nedd4 network that ameliorates alpha-synuclein toxicity in neurons. Science *342*, 979-983.

Temple, G., Gerhard, D.S., Rasooly, R., Feingold, E.A., Good, P.J., Robinson, C., Mandich, A., Derge, J.G., Lewis, J., Shoaf, D., *et al.* (2009). The completion of the Mammalian Gene Collection (MGC). Genome Res *19*, 2324-2333.

Ting, C.T., Tsaur, S.C., Sun, S., Browne, W.E., Chen, Y.C., Patel, N.H., and Wu, C.I. (2004). Gene duplication and speciation in Drosophila: evidence from the Odysseus locus. Proc Natl Acad Sci U S A *101*, 12232-12235.

Turner, B., Razick, S., Turinsky, A.L., Vlasblom, J., Crowdy, E.K., Cho, E., Morrison, K., Donaldson, I.M., and Wodak, S.J. (2010). iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database (Oxford) *2010*, baq023.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., *et al.* (2009). An empirical framework for binary interactome mapping. Nat Methods *6*, 83-90.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Veron, A.S., Kaufmann, K., and Bornberg-Bauer, E. (2006). Evidence of Interaction Network Evolution by Whole-Genome Duplications: A Case Study in MADS-Box Proteins. Mol Biol Evol *24*, 670-678.

Vidalain, P.-O., Boxem, M., Ge, H., Li, S., and Vidal, M. (2004). Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods 32*, 363-370.

Vingron, M., and Waterman, M.S. (1994). Sequence alignment and penalty choice. Review of concepts, case studies and implications. J Mol Biol *235*, 1-12.

Vo, T.V., Das, J., Meyer, M.J., Cordero, N.A., Akturk, N., Wei, X., Fair, B.J., Degatano, A.G., Fragoza, R., Liu, L.G., *et al.* (2016). A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human. Cell *164*, 310-323.

Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nat Protoc *2*, 1467-1478.

Wang, X., Wei, X., Thijssen, B., Das, J., Lipkin, S.M., and Yu, H. (2012). Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol *30*, 159-164.

Wei, X., Das, J., Fragoza, R., Liang, J., Bastos de Oliveira, F.M., Lee, H.R., Wang, X., Mort, M., Stenson, P.D., Cooper, D.N., *et al.* (2014). A massively parallel pipeline to clone DNA variants and examine molecular phenotypes of human disease mutations. PLoS Genet *10*, e1004819.

Wilkinson, M.G., Samuels, M., Takeda, T., Toone, W.M., Shieh, J.C., Toda, T., Millar, J.B., and Jones, N. (1996). The Atf1 transcription factor is a target for the Sty1 stress-activated MAP kinase pathway in fission yeast. Genes Dev *10*, 2289-2301.

Wilson-Grady, J.T., Villen, J., and Gygi, S.P. (2008). Phosphoproteome analysis of fission yeast. J Proteome Res *7*, 1088-1097.

Wood, V., Gwilliam, R., Rajandream, M.-A., M. Lyne1, R.L., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., *et al.* (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871-880.

Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan, B.W., Staines, D.M., Aslett, M., Lock, A., Bahler, J., Kersey, P.J., *et al.* (2012). PomBase: a comprehensive online resource for fission yeast. Nucleic Acids Res *40*, D695-699.

Woolcock, K.J., Stunnenberg, R., Gaidatzis, D., Hotz, H.R., Emmerth, S., Barraud, P., and Buhler, M. (2012). RNAi keeps Atf1-bound stress response genes in check at nuclear pores. Genes Dev *26*, 683-692.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., *et al.* (2008). High-quality binary protein interaction map of the yeast interactome network. Science *322*, 104-110.

Yu, H., Luscombe, N.M., Lu, H.X., Zhu, X., Xia, Y., Han, J.D., Bertin, N., Chung, S., Vidal, M., and Gerstein, M. (2004). Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res *14*, 1107-1118.

Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X., *et al.* (2011). Next-generation sequencing to generate interactome datasets. Nat Methods *8*, 478-480.