2011

EXPLORING FATC DOMAIN FUNCTION IN YEAST TRA1

Stephanie Marie Kvas

EXPLORING FATC DOMAIN FUNCTION IN YEAST TRA1

(Thesis Format: Monograph)

by

Stephanie Marie Kvas

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. Chris Brandl	Dr. Rob Hegele
Supervisory Committee	Dr. Joe Torchia
Dr. Megan Davey	Dr. Kathleen Hill
Dr. David Haniford	

The thesis by

Stephanie Marie Kvas

entitled:

Exploring FATC domain function in yeast Tra1

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

Date_____

_____
Chair of the Thesis Examination Board

Abstract

Tra1 is an essential yeast protein required for regulated transcription. Its human homolog TRRAP regulates factors important in oncogenesis. Mutation of leucine to alanine at position 3733 in the FATC domain ($tra1_{LA}$) results in growth phenotypes including sensitivity to ethanol. ***My aim was to examine genetic interactions of the FATC domain of Tra1 to define its cellular role.*** I screened for extragenic suppressors of the ethanol sensitivity caused by $tra1_{LA}$, identifying an opal mutation at tryptophan 165 of *NAM7* as a suppressor. Deleting *nam7*, *upf3*, or *nmd2* similarly suppressed $tra1_{LA}$, thereby linking Tra1 to nonsense mediated decay. I propose that Tra1 regulates transcription of genes also regulated by NMD. This work emphasizes the importance of NMD in gene regulation. Furthermore, the cross regulation between Tra1 and NMD suggests that mutations in the human NMD machinery may provide a mechanism to alter pathways influenced by TRRAP in human disease.

Keywords: Tra1, TRRAP, SAGA, NuA4, ASTRA, yeast, genetics, transcription, nonsense-mediated mRNA decay, Nam7, Upf3, Nmd2.

Acknowledgements

I want to especially thank my supervisor, Dr. Chris Brandl, for his valuable expertise, guidance, and for initiating this project. I thank him for always being there to answer questions – I have learned a lot from him. I thank Dr. Brandl for performing all of the tetrad dissections referred to in this thesis, and for thoroughly reviewing this thesis.

Thank you to my Thesis Advisory Committee, Dr. Megan Davey and Dr. David Haniford, for their help and advice regarding experiments. Thank you especially to Dr. Davey for reviewing this thesis.

I thank Dr. Greg Gloor for teaching me computer programming and Bioinformatics. His expertise in Bioinformatics was key in identifying the extragenic suppressor mutations referred to in this thesis.

I thank all Brandl lab members for their enthusiasm, encouragement, and for making the laboratory an enjoyable work atmosphere. Thank you to Julie Genereaux for her technical assistance and in helping with the phage display experiments. Thank you to Dom Dobransky, Stephen Hoke, Simon Lam, Aaron Simkovich and Heather Andrighetti for their help in the lab, and for all their support. Thank you to Simon for proof-reading this thesis.

Thank you to the University of Western Ontario for the Western Graduate Research Scholarship (WGRS) that funded my tuition. Thank you to NSERC for the Alexander Graham Bell Canada Graduate Scholarship.

I thank my family and friends for their constant support. I thank my parents (Mary and Peter), grandparents, and relatives for always encouraging me to do my best. I thank my siblings, Monica and Kevin, for constantly reminding me that anything is possible.

A very special thank you goes to my husband Owen. I thank him for his unwavering support, patience, and understanding throughout all aspects of this degree, and for always believing in me. I also thank him for assisting with preparing some of the figures in this thesis.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
ABI SOLiD	Applied Biosystems Sequencing by Oligonucleotide Ligation and Detection
ASTRA	ASsembly of Tel, Rvb and Atm-like kinase
ATP	Adenosine Triphosphate
bp	Base pair
BSA	Bovine Serum Albumin
BTI	Barley trypsin inhibitor
ChIP	Chromatin immunoprecipiation
DNA	Deoxyribonucleic acid
DSE	Downstream Sequence Element
DUB	De-UBiquitylation
EDTA	Ethylenediaminetetraacetic acid
FAT	FRAP, ATM, TRRAP
FATC	FRAP, ATM, TRRAP, C-terminus
GB	General Binding
GTFs	General Transcription Factors
g	gram
HAS	helicase-SANT-associated
HAT	Histone Acetylatransferase
HDAC	histone de-acetylase
HEAT	Huntintin, Elongation factor 3, A subunit of protein phosphatase 2A, and Tor1
hNuA4	Human NuA4
IPTG	isopropyl β-D-1-thiogalactopyranoside
KDa	Kilodalton
L	Litre
MDa	Megadalton
mRNA	Messenger RNA

µg	microgram
µL	microlitre
mg	milligram
mL	millilitre
NEB	New England Biolabs
NHEJ	Non-homologous end joining
NMD	Nonsense Mediated mRNA Decay
NuA4	Nucleosomal Acetylatransferase of histone H4
PCR	Polymerase Chain Reaction
PIC	Pre-initiation complex
PI3K	Phosphatidylinositol-3-kinase
PIKK	phosphatidylinositol 3-kinase-related kinase
PMSF	phenylmethylsulfonyl fluoride
PVDF	Polyvinylidene fluoride
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
RSC	Remodels the Structure of Chromatin
SAGA	Spt3-Ada-Gcn5-acetyltransferase
SAM	Sequence Alignment/Map
SANT	SWI3-ADA2-N-CoR-TFIIIB
SD	Synthetic Dropout
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGD	*Saccharomyces cerevisiae* Genome Database
SLIK	SAGA-like complex
STAGA	Spt3-Taf9-Ada-Gcn5-acetyltransferase
SWI/SNF	SWItch/Sucrose NonFermentable
TAF	TATA-binding protein associated factors
TBP	TATA-binding protein
TBS	Tris buffered saline
TE	Tris-EDTA

TFIID	Transcription Factor II D
TRRAP	TRansformation/TRanscription domain-Associated Protein
UAS	Upstream Activating Sequence
YPD	Yeast Peptone Dextrose

Chapter1: Introduction

1. Prelude

The 433 kDa protein Tra1 is a transcriptional regulator found in several protein complexes in *S. cerevisiae*. The human homolog of Tra1, TRRAP, regulates transcription factors required for oncogenesis. Tra1's role in transcription, presence in multiple transcriptional regulatory complexes, and membership in the phosphatidylinositol 3-kinase-related kinase (PIKK) family of proteins, suggests that Tra1 is a key player in eukaryotic biology. Interestingly, unlike other PIKK family members, Tra1 lacks kinase activity. The goal of the Brandl laboratory is to identify the cellular functions and molecular interactions of Tra1. ***The goal of my thesis is to examine genetic interactions of the FATC domain of Tra1 to gain more information on Tra1's cellular role.***

This introduction will highlight several topics related to gene expression with a focus on Tra1. These topics include: 1) analysis of transcriptional regulation in eukaryotes using yeast as a model system, 2) the three transcriptional regulatory complexes that contain Tra1; SAGA, NuA4, and ASTRA, 3) the PIKK family of proteins, of which Tra1 is a member, and 4) mRNA processing. Nonsense-mediated mRNA Decay (NMD) will also be discussed in some detail since the experiments performed in this thesis suggest that Tra1 may have a role in this process through its interactions with the protein Nam7. The theme of crosstalk between complexes that include Tra1 will be evident.

2. Eukaryotic Transcription

2.1 Yeast as a model organism

Saccharomyces cerevisiae (budding yeast) is a single-celled eukaryote. Many genes and cellular processes are conserved from yeast to humans, making yeast an excellent organism for understanding cellular processes, such as transcription, DNA replication, and protein trafficking in higher eukaryotes. In addition, since yeast stably exists as a haploid or diploid, it is well suited for classical genetics studies. The entire yeast genome was sequenced in 1996, making studies on yeast even simpler (Goffeau et al. 1996). Improved mass spectrometry techniques resulted in a flood of proteins being characterized, thereby identifying some of the key players involved in important cellular processes. Increased knowledge of the cellular processes in yeast will aid in our understanding of these same cellular processes in more complex eukaryotes, such as humans.

2.2 Transcriptional Regulation

Transcription is critical for the growth and development of all eukaryotes, whereby a myriad of protein complexes whose co-ordinated activities are required (Cairns. 2009). These protein complexes regulate and control transcription so that gene expression may successfully occur. Diseases, including certain cancers, are a result of altered gene expression often caused by transcriptional mis-regulation (Villard. 2004).

Transcription requires RNA polymerase II (RNAPII), the General Transcription Factors (GTFs), and gene specific regulators. The GTFs bind at the promoter through the action of the TATA-binding protein (TBP) and are directly involved in the recruitment of RNAPII (Figure 1). The association of GTFs and RNAPII form the pre-initiation

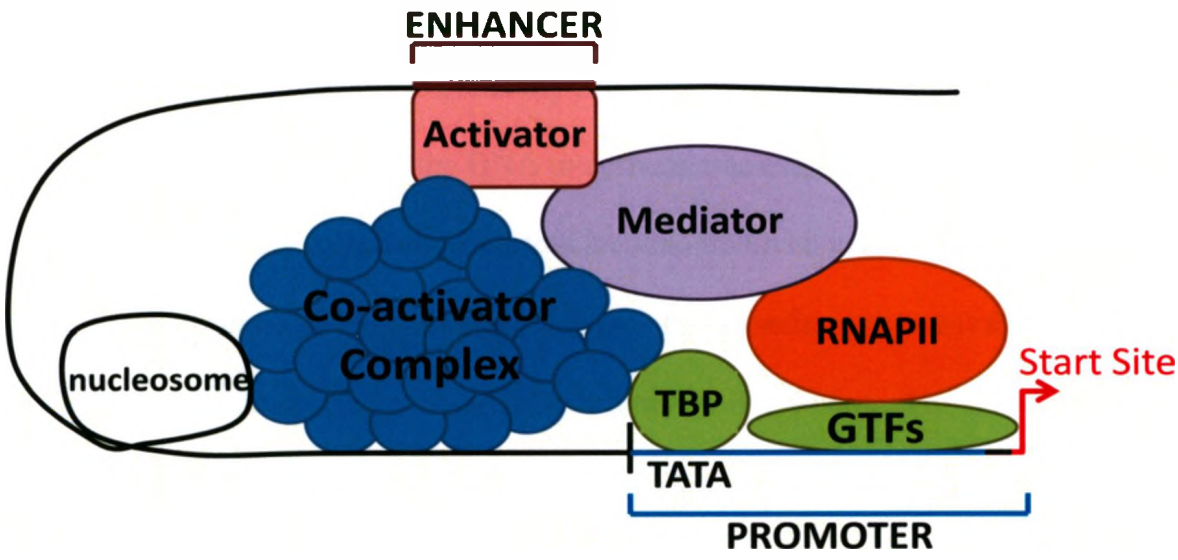

Figure 1: Eukaryotic transcription is highly regulated. Co-activator complexes remodel chromatin and direct activators to DNA sequences upstream of the promoter. Mediator is a co-activator complex mediating protein interactions between the activator and transcriptional machinery. General Transcription Factors (GTFs) bind at the promoter through the action of the TATA-binding protein (TBP) and recruit RNA polymerase II (RNAPII) to the transcription start site. The association of GTFs, TBP, and RNAPII form the pre-initiation complex (PIC) that allows transcription to begin.

complex (PIC) that allows transcription to begin (Bhaumik et al. 2004, Green. 2005, Martinez. 2002). Activators are generally sequence specific DNA binding proteins that bind to a regulatory region on the DNA to activate transcription. An additional set of molecules generically called co-activators are also involved in regulating transcription (Figure 1). This class of molecules can enhance (or sometimes repress) the activity of activator proteins and/or help the activators function in the context of chromatin. Transcriptional co-activator complexes are large protein machines containing multiple subunits. These co-activator complexes are often recruited to promoters through one of their subunits, providing activator proteins access to DNA (Martinez. 2002, Naar et al. 2001).

2.3 Chromatin remodeling

Activator proteins must access regulatory sequences on DNA so that transcription can occur. Highly conserved histone proteins (H1, H2A, H2B, H3, and H4) function as building blocks to compact eukaryotic DNA. These histone proteins form octamers and the DNA wraps around the histones to form repeating nucleosomes, the basic unit of DNA packaging. These nucleosomes then fold into higher order structures to form highly compact chromatin (Strahl and Allis. 2000). In yeast, a regulatory region is sometimes "open", facilitating activator binding to the promoter (Cairns. 2009). These constitutively open promoters are often linked to housekeeping genes. However, if a regulatory region and promoter is "covered" (located in a region of compact chromatin), chromatin remodeling is a necessary step to uncover the regulatory sequence and promoter so that activators can bind (Cairns. 2009). Chromatin-modifying enzymes are required for this process (Narlikar et al. 2002).

Generally there are two ways of altering chromatin. Chromatin remodeling complexes can change the position of DNA wrapped around histones, using ATP hydrolysis to disrupt histone-DNA interactions, thereby loosening the DNA and making it more accessible to regulatory proteins (Strahl and Allis. 2000). Yeast and mammalian SWI/SNF (SWItch/Sucrose NonFermentable) and RSC (Remodels the Structure of Chromatin) complexes are examples of chromatin remodeling complexes that alter chromatin structure in this manner (Strahl and Allis. 2000). Chromatin can also be altered by chromatin modifying enzymes that covalently modify histones in chromatin through lysine acetylation, lysine and arginine methylation, serine phosphorylation, or ubiquitination. Modifications occur at specific residues, usually on N-terminal histone tails (extensions from the histone proteins). These tails provide a surface for interactions with other proteins (Martinez. 2002, Strahl and Allis. 2000, Daniel and Grant. 2007). The best studied form of chromatin modification is histone acetylation of lysine residues catalyzed by histone acetyltransferases (HATs). The simplest interpretation for the role of histone acetylation is to neutralize the charge on lysine, which weakens the interaction between DNA and histones, thereby allowing the chromatin to loosen and transcription factors greater access to DNA (Martinez. 2002, Allard et al. 1999). Histone methylation, at lysines or arginines, is generally involved in repressing transcription through condensing chromatin; however, methylation can also loosen chromatin to activate transcription (Kouzarides. 2002). Histone phosphorylation is generally involved in activating transcription and is thought to initiate chromosome condensation during mitosis (Strahl and Allis. 2000). Histone ubiquitination has been linked to both gene activation and repression (Weake and Workman. 2008).

Different histone modifications can create specific patterns to attract protein domains present in various regulators; for example, bromodomains recognize acetylated lysines and chromodomains recognize methylated lysines (Auger et al. 2008, Keogh et al. 2006). The histone code hypothesis predicts that covalent histone modifications act as signals to recruit specific protein complexes to DNA to alter transcription and other downstream cellular events (Strahl and Allis. 2000). Therefore, transcriptional co-regulatory complexes contribute to the histone code through modifying chromatin (Keogh et al. 2006).

3. Transcriptional co-activators containing Tra1

Two large complexes found in yeast, SAGA (Spt-Ada-Gcn5-acetyltransferase) and NuA4 (Nucleosomal Acetylatransferase of histone H4), are transcriptional co-activator complexes containing multiple subunits involved in altering chromatin and regulating transcription (Allard et al. 1999, Grant et al. 1997). Homologous complexes of both SAGA and NuA4 are found in humans, STAGA (Spt3-Taf9-Ada-Gcn5-acetyltransferase) and hNuA4 (human NuA4) (Auger et al. 2008, Park et al. 2001).

3.1 SAGA

SAGA is a 2.0 MDa transcriptional co-activator complex containing 19 proteins that regulate a number of cellular processes, most notably for this thesis, transcription (Bhaumik et al. 2004, Baker and Grant. 2007). Overall, SAGA is responsible for the transcription of about ten percent of yeast genes, most of which are involved in stress response (Lee et al. 2000, Huisinga and Pugh. 2004). Electron microscopy revealed that yeast SAGA is divided into several modular domains (Figure 2; Wu et al. 2004).

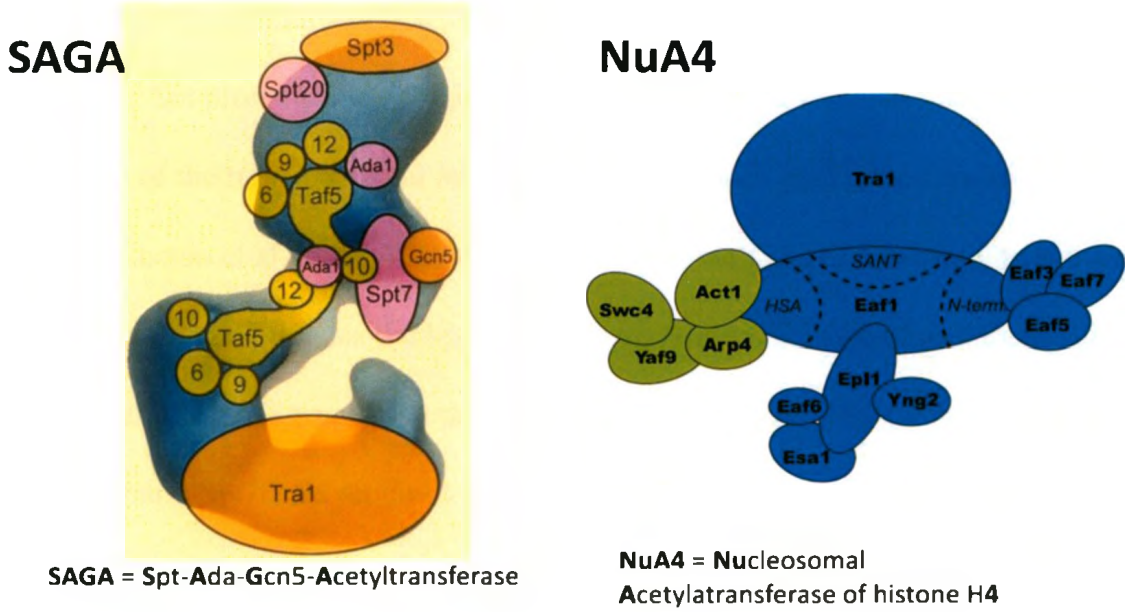

Figure 2: Tra1 is a member of SAGA and NuA4 modular protein complexes. SAGA, determined by electron microscopy, is shown on the left. Tra1 sits at the periphery of the complex, made up of modular domains (Figure from: Wu, et al. 2004) NuA4, structure not yet determined, is also made up of modular domains. Tra1 is a modular domain and connects to Eaf1, a subunit required for assembling the NuA4 complex (Figure from: Lu, et al. 2009).

One group of molecules within SAGA is the Spt proteins (Spt3, Spt7, Spt8, Spt20). The Spt proteins were originally identified by Fred Winston's laboratory as suppressors of the transcriptional inhibitory effects caused by Ty transposable elements in yeast (Winston et al. 1984, Winston et al. 1987). One class of Spt proteins is the chromatin class and includes the histone proteins. Another class includes those Spt proteins found in SAGA and also the TATA binding protein (TBP), Spt15 (Hahn et al. 1989). Spt7 and Spt20 are required for the structural integrity of SAGA (Martinez. 2002, Wu et al. 2004). Deletions of *spt7* or *spt20* result in severe phenotypes in yeast (Wu et al. 2004). Spt3 and Spt8 make up the TBP (TATA-binding protein) interaction module because they facilitate an interaction between SAGA and TBP (Martinez. 2002, Wu et al. 2004). Deletion of either s*pt3* or *spt8* is less severe than that of deletion of s*pt7* or s*pt20*. There is another version of SAGA known as the SLIK (SAGA-like) complex (Grant et al. 1998). SLIK contains the same subunits as SAGA, except for Spt8. Therefore, Spt8 is the only subunit completely unique to SAGA. The SLIK complex, like SAGA, is also involved in histone acetylation and transcription (Baker and Grant. 2007).

The Ada proteins (Ada2, Ngg1/Ada3, Gcn5/Ada4) form the HAT module of SAGA (Berger et al. 1992, Brandl et al. 1993, Brownell et al. 1996). The *ADA* genes were first identified in the Guarente laboratory as suppressors of the toxicity caused by overexpression of the strong transcriptional activator VP16 (Berger et al. 1992). Ada1 is also part of this group, though it shares more functional similarity with Spt7 and Spt20 since it is required for the structural integrity of SAGA (Martinez. 2002, Wu et al. 2004, Burgess et al. 2010). Subsequently, function for the Ada proteins in both the activation and repression of transcription were found (Jacobson and Pillus. 2009). A landmark

finding in the transcriptional field was the discovery that Gcn5 possesses HAT activity (Brownell et al. 1996). The Workman laboratory identified the HAT module within SAGA; Gcn5 being the catalytic subunit and its acetylation of histones H2B and H3 being regulated by Ada2 and Ngg1 (Grant et al. 1997). Once SAGA is recruited to the promoter (via Tra1, see below) histone acetylation takes place in a targeted fashion, thereby altering the chromatin landscape for transcription factor recruitment (Martinez. 2002). Interestingly, the Ada proteins also seem to function independently of SAGA, a point made evident from the fact that a double deletion of both *gcn5Δ* and *spt7Δ* is more severe than *spt7Δ* alone (Wu and Winston. 2002). This agrees with the finding that Gcn5 also acts in global untargeted acetylation (Imoberdorf et al. 2006). Sgf29 might also be considered within the Ada group. It is a poorly studied subunit of the SAGA complex, but it directly interacts with Ngg1 and shares some of the common phenotypes seen in the HAT module (Kurabe et al. 2007).

SAGA also possesses histone de-ubiquitylation activity. Histone ubiquitylation is linked to both gene activation and repression. The DUB (De-UBiquitylation) module in SAGA consists of: Ubp8, Sgf73, Sgf11, and Sus1. Ubp8 is the catalytic subunit that removes ubiquitin from the H2B C-terminal tail, promoting methylation and causing an increase in active chromatin (Baker and Grant. 2007). Sgf73 physically connects the DUB module with the rest of the SAGA complex (Koutelou et al. 2010). The Sgf73 and Sus1 proteins also link SAGA to the mRNA export machinery (Rodriguez-Navarro. 2009, Koutelou et al. 2010).

SAGA contains many of the TATA-binding protein associated factors (TAFs), also found within the GTF TFIID (Transcription Factor II D). *TAF9*, *TAF10*, and *TAF12*

are all essential genes, perhaps due to their role within both TFIID and SAGA. There is both a structural and functional conservation between SAGA and TFIID (Martinez. 2002). However, whereas TFIID more globally regulates gene expression, SAGA mainly regulates gene expression in response to stress (Huisinga and Pugh. 2004).

The final module of SAGA is composed of the single molecule Tra1. Tra1 will be discussed in greater detail below. In this context it is important to note that Tra1 interacts directly with transcriptional activators (Baker and Grant. 2007, Brown et al. 2001, Fishburn et al. 2005, Reeves and Hahn. 2005). As such it is the module that provides one mechanism to recruit SAGA to promoters.

In addition to its role in transcription activation, SAGA functions in other cellular processes such as in response to DNA damage, and nucleotide excision repair (Baker and Grant. 2007). SAGA, through the Gcn5 subunit, is required for normal cell cycle progression. Gcn5 mutants accumulate at the G2/M phase of the cell cycle (Burgess et al. 2010). Although many SAGA members have been identified, it is still unclear how all of them interact, coordinate their enzymatic activities, and regulate transcription (Koutelou et al. 2010).

3.2 NuA4

NuA4 is a 1.3 MDa transcriptional co-activator complex that regulates multiple cellular processes, including transcription (Allard et al. 1999). NuA4 contains 13 subunits which make up 5 modular domains. Figure 2 depicts NuA4 and its modular domains (Lu et al. 2009).

NuA4 exists as both a large 13 subunit complex, and as a smaller complex, known as piccolo NuA4. The piccolo NuA4 complex forms one module in NuA4 and consists of

Esa1, Yng2, Eaf6, and Epl1 (Lu et al. 2009). NuA4 acetylates histones H2A, H4, and the histone variant Htz1 through the Esa1 catalytic subunit (human homolog Tip60) (Lu et al. 2009). Allard et al. demonstrated that Esa1 can acetylate histones independently, however, it can only stimulate transcription in the context of NuA4 (Allard et al. 1999). Esa1 is also essential for cell cycle progression in yeast (Allard et al. 1999). NuA4 is the only essential HAT complex in yeast, perhaps because it acts in multiple pathways (Lu et al. 2009).

Tra1 acts to target NuA4 to transcriptional activators, a role it also has within the SAGA complex. NuA4 is recruited by Tra1 to the promoter region of many genes, such as the ribosomal protein genes, to regulate transcription (Auger et al. 2008).

The Eaf1 subunit is exclusively found within NuA4 (Auger et al. 2008), functioning as a scaffold module at the centre of the complex. Eaf1 physically interacts with and coordinates assembly of the other four modules into NuA4 (Auger et al. 2008, Mitchell et al. 2008). Eaf1 has various domains that connect it to the other modules. For example, the SANT (SWI3-ADA2-N-CoR-TFIIIB) domain of Eaf1 binds to Tra1 (Auger et al. 2008, Lu et al. 2009). The N-terminal region of Eaf1 associates with the Eaf3, Eaf5 and Eaf7 module, whose function is unclear (Lu et al. 2009). Eaf5 directly interacts with Eaf1, allowing Eaf7 and Eaf3 subunits to associate with the NuA4 complex. Eaf3, Eaf4, and Eaf5 also exist as an independent trimer complex (Auger et al. 2008).

Another module consists of Swc4, Act1, Yaf9, and Arp4, which are shared subunits between NuA4 and SWR1, another chromatin remodeling complex involved in regulating transcription and DNA repair (Lu et al. 2009). Arp4 directly interacts with Eaf1 through the HAS (helicase-SANT-associated) domain to connect this module to

NuA4 (Lu et al. 2009). SWR1 and NuA4 subunits show strong genetic interactions, which may reflect the cross-talk between these two complexes (Auger et al. 2008). For example, NuA4 acetylates H4, allowing SWR1 to recognize this acetylation and insert histone variant Htz1 into chromatin. NuA4 subsequently acetylates Htz1, so these two complexes work together (Keogh et al. 2006, Lu et al. 2009). Interestingly, the human homolog of NuA4, hNuA4, is an exact physical fusion of SWR1 and NuA4 so only a single recruitment takes place (Auger et al. 2008).

In addition to its roles in histone acetylation and transcriptional regulation, NuA4 is also involved in chromosome segregation, DNA replication, DNA repair, and cell cycle progression (Lu et al. 2009). Distinct patterns of histone H4 acetylation by NuA4 are essential in multiple pathways of double-strand break repair, and NuA4 mutants are sensitive to genotoxic agents causing double strand breaks (Bird et al. 2002). Esa1 is needed to acetylate H4 in order for NHEJ (non-homologous end joining), a method of double strand break repair in which chromosome ends are joined together, to take place (Bird et al. 2002).

Human NuA4 regulates gene expression by numerous transcriptional regulators, including E2F1, Myc, and p53 (Lu et al. 2009). Human NuA4, like NuA4 and SWR1, is involved in DNA repair. In the context of these roles, hNuA4 functions as a tumour suppressor; interestingly, its expression is reduced in multiple cancer types (Lu et al. 2009).

3.3 ASTRA

The ASTRA (ASsembly of Tel, Rvb and Atm-like kinase) complex is a recently discovered chromatin remodeling protein complex potentially involved in regulating

transcription (Shevchenko et al. 2008). All of the components of ASTRA are essential for yeast viability. ASTRA is composed of the Rvb1/Rvb2 heterodimer, an essential helicase also present in two other chromatin remodeling complexes, SWR1 and INO80 (Jonsson et al. 2004). Rvb1 and Rvb2 are required for the assembly and function of the INO80 chromatin remodeling complex (Jonsson et al. 2004). In humans, Rvb1/Rvb2 associate with the transcription factor and oncoprotein Myc, and mutations in Rvb1/Rvb2 prevent cellular transformation by Myc (Jonsson et al. 2004).

Three less characterized proteins, Tti1, Tti2, and Asa1, are also part of ASTRA (Shevchenko et al. 2008). Although the functions of Tti1 and Tti2 are unknown, they interact with the telomere binding protein Tel2, another component of the ASTRA complex. Tel2 is an essential protein, originally identified for its involvement in regulating telomere length and positioning in yeast, however, it does not do this in humans (Runge and Zakian. 1996, Takai et al. 2007). Human homologs of Tti1 and Tti2 exist with Tel2 as a "Triple T" complex required for DNA damage response (Hurov et al. 2010). Intriguingly, Tel2 physically interacts with and is essential for the stability of all PI3K-like (PIKK) family members (Takai et al. 2007, Kanoh and Yanagida. 2007, Horejsi et al. 2010). Tel2 is phosphorylated by CK2 and this phosphorylation is required for mTOR and SMG1 stability in mouse cells (Horejsi et al. 2010). Since Tel2 is essential for the stability of all the PIKKs, it is likely an important regulator or coordinator of PIKK activity. Tel2 preferentially binds to newly synthesized PIKKs (Takai et al. 2007). Additionally, the interaction between Tel2 and Hsp90 is required for Tel2 to interact with certain PIKKs such as ATM, ATR, mTOR and DNA-PKcs (Takai et al. 2010). Furthermore, the interaction between Tel2, Tti1 and Tti2 is important for Tel2 function

(Takai et al. 2010).

Tra1 is also present in ASTRA. The role of Tra1 in the context of ASTRA is unclear. Perhaps a clue comes from the fact that a triple alanine mutation ($tra1_{SRR}$) at the C-terminus of Tra1 results in generation dependent telomere shortening (Mutiu et al. 2007). It is interesting to speculate that this telomere shortening results from Tra1's role within ASTRA.

4. Tra1/TRRAP

Tra1 (TRanscription-Associated protein 1) is a 433 kDa essential protein involved in co-activating transcription through directly binding to acidic activators such as Gcn4, VP16, Gal4 and Hap4 (Saleh et al. 1998, Brown et al. 2001). In humans, the homologous TRRAP (TRansformation/TRanscription domain-Associated Protein) is required for transcriptional regulation and functions similarly to Tra1 to recruit STAGA and large multiprotein HAT complexes to gene promoters (Park et al. 2001). TRRAP regulates transcription factors, such as Myc, and p53, which are important in oncogenesis (McMahon et al. 1998, Murr et al. 2007). In addition, TRRAP is required for early embryonic development and cell cycle progression (Herceg et al. 2001).

TRRAP was first identified as a nuclear protein that interacted with the N-terminus of Myc and was an essential cofactor for Myc cellular transformation (McMahon et al. 1998). Tra1 was then identified as the yeast homolog of TRRAP, containing approximately 55-60 % similarity (McMahon et al. 1998). Yeast Tra1 was first shown to associate with SAGA components in 1998 by the Brandl laboratory (Saleh et al. 1998). Shortly afterwards, the Workman laboratory also identified Tra1 as a component of both

SAGA and SLIK (Grant et al. 1998). Tra1 was purified as a component of the NuA4 complex by the Côté laboratory in 1999 (Allard et al. 1999).

Tra1 begins the chain of events leading to transcriptional activation of several genes. For example, Tra1 directly binds to the acidic activator Gal4, which is the DNA binding transcription factor responsible for activating transcription of the *GAL* genes (Bhaumik and Green. 2001). The interaction between Tra1 and Gal4 recruits SAGA to the upstream activating sequence (UAS), where Gal4 binds (Bhaumik et al. 2004, Bhaumik and Green. 2001). SAGA then recruits the Mediator complex, which recruits the GTFs and RNAPII to the promoter to initiate transcription (Bhaumik et al. 2004).

4.1 Structure of Tra1/TRRAP

Tra1/TRRAP are part of the PI3K-like (PIKK) family of proteins. Figure 3 illustrates the C-terminal domains of Tra1. Members of this family are high molecular mass proteins (300-500 kDa) that are involved in stress response, transcription, DNA repair, and cell cycle control (Keith and Schreiber. 1995, Abraham. 2004). The amino terminal domain of PIKK proteins is composed of 40-50 HEAT (Huntintin, Elongation factor 3, A subunit of protein phosphatase 2A, and Tor1) repeats, which may be protein interaction sites. The catalytic domain of PIKK family members is a 300 amino acid phosphatidylinositol-3-kinase (PI3K) domain, which transfers phosphate from ATP to proteins, rather than lipids (Keith and Schreiber. 1995, Abraham. 2004). The PI3K domain is flanked by a 500 amino acid FAT (FRAP, ATM, TRRAP) domain upstream, and directly followed by a 35 amino acid FATC (FRAP, ATM, TRRAP, C-terminus) domain downstream at the C-terminus (Park et al. 2001, Abraham. 2004). The FAT and FATC domains, and the large size of these molecules, are unique features of the PIKK

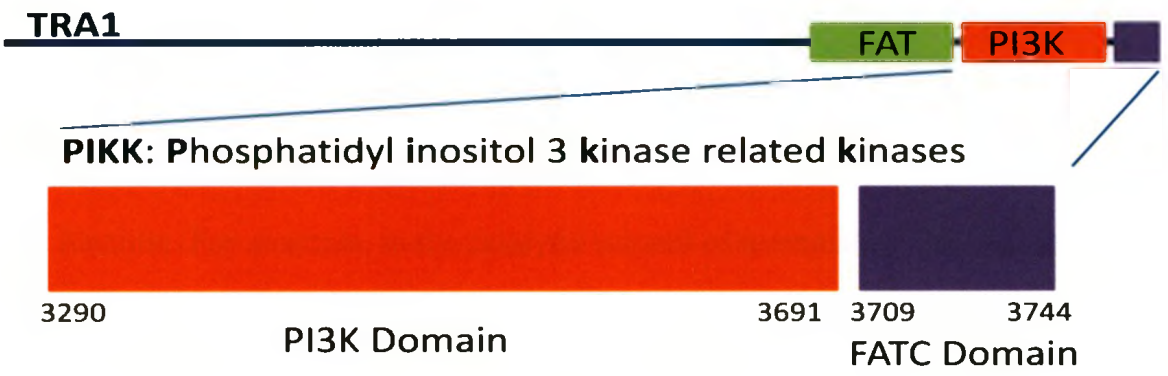

Figure 3: The C-terminal domains of Tra1. Tra1 belongs to the PIKK family of proteins. At the C-terminus, members of this protein family contain a FAT domain, followed by a PI3K domain and FATC domain. The FATC domain normally regulates the kinase activity of the PI3K domain. However, the PI3K domain of Tra1 does not possess kinase activity.

family (Bosotti et al. 2000).

Since PIKK family members all have common domains and are involved in similar processes (stress response), it seems likely that their structure plays a key role in their function. For example, in the catalytic subunit of mammalian DNA-dependent protein kinase (DNA-PKcs), the C-terminal domain of the protein loops back to interact with N-terminal sequences (Spagnolo et al. 2006). Furthermore, the FATC domain in other members of the PIKK family, such as mTOR and SMG1, is required for regulating the kinase activity of the PI3K domain (Takahashi et al. 2000, Morita et al. 2007). The FATC domain of DNA-PKcs is also required for the kinase activity of PI3K domain (Priestley et al. 1998, Beamish et al. 2000). Interestingly, the PI3K domain of Tra1/TRRAP is missing the catalytic residues needed for kinase activity (Saleh et al. 1998, Bosotti et al. 2000). However, the conservation of the PI3K/FATC domains across species implies that the C-terminus of Tra1 is important for function.

4.2 Mutations of the C-terminus of Tra1

As mentioned above, the FATC domain of other PIKK family members is required for the kinase activity of PI3K domain. However, Tra1/TRRAP does not possess kinase activity (Saleh et al. 1998). Studies of various mutations in the PI3K and FATC domains of Tra1 have shown that these two domains are important for function (Mutiu et al. 2007, Hoke et al. 2010) Other studies have shown that the C-terminal domains of some PIKK proteins are involved in protein-protein interactions. For example, the FATC domain of yeast Mec1 (human homolog ATR) interacts with Rfa1, thereby recruiting Mec1 to DNA damage sites (Nakada et al. 2005). Another example is the FATC domain of human ATM is required for the HAT activity and interaction between Tip60 and

ATM, which is activated by DNA damage (Sun et al. 2005). Mutations in the FATC domain of ATM abolish the interaction between ATM and Tip60, causing defects in the DNA damage response pathway (Sun et al. 2005). Likewise, the C-terminus of Tra1/TRRAP may also be involved in protein-protein interactions. One study found that when the PI3K or FATC domain of TRRAP was deleted, it was no longer able to bind to Gcn5 and recruit HAT activity (Park et al. 2001). Therefore, the C-terminus of TRRAP is important for the interaction and HAT activity of Gcn5 to occur (Park et al. 2001). Hence, despite the fact that Tra1/TRRAP possess no kinase activity, there is evidence that the C-terminal domains are important regulators and may be involved in protein-protein interactions.

A study in the Brandl laboratory performed a targeted mutagenesis in regions of Tra1 that may be relevant for function using triple alanine scanning mutations. A PI3K domain mutant of Tra1 ($tra1_{SRR}$) exhibited growth phenotypes such as sensitivity to ethanol and media lacking phosphate (Mutiu et al. 2007). The $tra1_{SRR}$ allele also reduced β-galactosidase activity at the *PHO5* promoter. Some of the phenotypes exhibited by $tra1_{SRR}$ are characteristic of SAGA and NuA4 mutants. However, the $tra1_{SRR}$ allele also resulted in shortened telomeres, which may be due to Tra1's role in ASTRA (Mutiu et al. 2007).

Analysis of mutations in the FATC domain of Tra1 has shown that the FATC domain is required for function (Hoke et al. 2010). The Brandl laboratory targeted residues that were conserved in Tra1 and among the PIKK family members. They found that within the FATC domain of Tra1, a leucine to alanine change at amino acid position 3733 ($tra1_{LA}$) renders yeast sensitive to various conditions of cellular stress, including

ethanol and temperature extremes (Hoke et al. 2010). In addition, $tra1_{LA}$ changes the cellular transcription profiles relative to the wild-type allele and results in decreased protein levels of Tra1 (Hoke et al. 2010). Like $tra1_{SRR}$, $tra1_{LA}$ showed many severe phenotypes, some characteristic of SAGA and NuA4 mutants. However, $tra1_{LA}$ did not show shortened telomeres like $tra1_{SRR}$. Interestingly, a corresponding mutation to $tra1_{LA}$ in *Drosophila* SMG1 results in reduced kinase activity of PI3K (Morita et al. 2007). Therefore, the $tra1_{LA}$ allele was of great interest in elucidating the FATC domain function of Tra1.

A phenylalanine to alanine change at position 3744 at the end of the FATC domain of Tra1 ($tra1_{FA}$), also resulted in severe phenotypes and illustrated the importance of the C-terminus of Tra1 (Hoke et al. 2010). The $tra1_{FA}$ mutation showed severe phenotypes and reduced viability (Hoke et al. 2010). Like $tra1_{LA}$, $tra1_{FA}$ resulted in transcriptional changes and growth defects. Furthermore, the Brandl laboratory showed that positioning of the C-terminus of Tra1 is important for function, since addition of a glycine to the terminal amino acid residue in Tra1 does not support viability in yeast (Hoke et al. 2010). Although the role of the FATC domain in Tra1 has not been characterized, evidence clearly suggests that it is critical for the function of the protein.

4.3 Studies of Tra1 in *Schizosaccharomyces pombe*

Due to its large size, researchers originally thought that Tra1 may be a scaffolding protein, or required for the assembly of SAGA. However, when the 3D structure of SAGA became available through electron microscopy (Wu et al. 2004), Tra1 was revealed at the periphery of the complex (see Figure 2). Hence, this positioning made it unlikely that Tra1 was a scaffolding protein. Furthermore, recent studies of Tra1 in

Schizosaccharomyces pombe (fission yeast) show that Tra1 is not needed for the assembly of SAGA (Helmlinger et al. 2011). In *S. pombe*, there are two versions of Tra1, Tra1 and Tra2. Tra1 is associated with SAGA and ASTRA and is non-essential, whereas Tra2 is exclusively associated with NuA4 and is essential (Helmlinger et al. 2011). Interestingly, in a *tra1Δ* in *S. pombe*, SAGA is still recruited to some promoters, but not others (Helmlinger et al. 2011).

Studies of Tra1 in *S. pombe* also suggest that Tra1's function within ASTRA may be similar to its role in SAGA, since Tra1 co-purifies with SAGA and several ASTRA subunits, whereas Tra2 co-purifies exclusively with NuA4 (Helmlinger et al. 2011). Genetic studies indicate that expression of many stress response genes that were altered in *tra1Δ*, were similarly altered in *spt7Δ* and *ada1Δ* (Helmlinger et al. 2011). More genes that were also affected by *tra1Δ*, were not affected by *spt7Δ* or *ada1Δ*, suggesting that these genes might be controlled by ASTRA or another complex (Helmlinger et al. 2011). This result emphasizes that Tra1's role in the context of all the different complexes in which it is found must be considered. In *S. cerevisiae* it is possible that Tra1 is involved in coordinating the activities of these three protein complexes, or perhaps Tra1 has additional roles.

5. PIKK Protein Family

Important cellular regulators such as mammalian ATM, ATR, mTOR, SMG1, and DNA-PKcs are part of the PIKK family of proteins (Abraham. 2004). PIKKs are relevant to human health as they are involved in many important cellular processes. ATM and ATR are involved in cell-cycle checkpoint pathways activated by DNA damage or DNA

replication stress. Individuals with mutations in ATM can develop a disease called ataxia telangiectasia, which can also increase the individual's risk of developing breast cancer (Ahmed and Rahman. 2006). Another PIKK protein, mTOR, coordinates protein synthesis, cell growth and cell proliferation with the availability of nutrients and mitogenic growth factors. DNA-PKcs regulates the repair of DNA double-strand breaks by the NHEJ pathway. SMG1 is involved in an RNA surveillance mechanism called nonsense-mediated mRNA decay (NMD), which will be discussed in the next section.

The crystal structure of DNA-PKcs was recently solved (Sibanda et al. 2010). The N-terminal portion of the protein contains HEAT repeats, which are helix-turn-helix repeats allowing the protein to bend. The overall form of DNA-PKcs is circular in structure, with the C-terminal domains, including the FATC domain, sitting at the top of the protein like a crown (Sibanda et al. 2010). The structure is flexible, allowing for DNA-double-strand-break repair to occur through the ring (Sibanda et al. 2010). Since members of the PIKK family are highly conserved, a similar architecture may be shared. The NMR structure of isolated FATC domain of yeast Tor1 has also been solved (Dames et al. 2005). In the oxidized form, the FATC domain is an alpha-helical structure with a loop at the C-terminus held together by two conserved cysteine residues. In the reduced form of Tor1, there is more flexibility in this loop structure since the disulfide bond is broken between the two cysteines (Dames et al. 2005). However, the C-terminus of Tra1 lacks these two conserved cysteines, therefore it is possible that the FATC domain structure is always flexible. Despite Tra1/TRRAP's importance and relevance to disease, little structure/function information is available.

6. mRNA Processing and Nonsense Mediated mRNA Decay (NMD)

During transcription, as the RNA is transcribed by RNAPII, it is subject to modifications before it leaves the nucleus and ultimately translation in the cytoplasm. Two modifications that mRNA undergoes is the addition of a 5' methylguanosine cap to protect the 5' end, and 3' polyadenylation, to protect the 3' end (Alberts et al. 2010). These modifications are also key signals for the RNA export machinery as they identify intact mRNAs to be transported through nuclear pores (Alberts et al. 2010). Splicing and editing of the mRNA also occurs before it is transported out of the nucleus (Alberts et al. 2010).

The importance of mRNA turnover in gene regulation has become increasingly apparent (Gonzalez et al. 2001). Messenger RNAs can be translated or degraded depending on the cellular situation. Some mutations in mRNA can lead to the insertion of a premature stop codon, also known as a nonsense mutation (Mitchell and Tollervey. 2003). If translated, nonsense mutations can result in truncated proteins that may be toxic to the cell. As well as being introduced through errors in transcription by RNAPII, premature stop codons can result from failure to remove introns or inaccurate translational starts (Wilusz et al. 2001). To reduce the number of potentially toxic truncated proteins, all eukaryotes have a process known as nonsense-mediated mRNA decay (NMD), responsible for degrading mRNAs containing premature stop codons (Wilusz et al. 2001). In yeast, translation termination is the first step in triggering NMD. When a stop codon is read by the ribosome during the translation of an mRNA, the ribosome signals to two translation termination factors, eRF1 and eRF3 (Mitchell and Tollervey. 2003). The next step is to determine whether the translation termination was

premature. In mammalian cells, where splicing of introns occurs, stop codons over 50 nucleotides upstream of the last exon-exon junction are recognized as premature and get targeted for NMD (Wilusz et al. 2001). Since yeast generally do not contain introns, the protein Hrp1 recognizes specific sequences in mRNA called downstream sequence elements (DSE) (Mitchell and Tollervey. 2003). One or more DSE sequences (having the consensus TGYYGATGYYYY) may be present within one mRNA (Zhang et al. 1995). eRF1 releases the polypeptide chain when a stop codon is read; whether a stop codon is recognized as premature depends on whether Hrp1 is bound to the DSE (Gonzalez et al. 2001). If the stop codon (correct or not) occurs downstream of the DSE sequence, Hrp1 cannot bind to the DSE sequence and the stop codon is believed to be the correct one. If Hrp1 is bound to the DSE and a stop codon is read by the ribosome, a surveillance complex recognizes the mRNA as containing a premature stop codon, and NMD is triggered. eRF1 then interacts with the NMD surveillance complex (Wilusz et al. 2001). Figure 4 illustrates the various situations in which mRNA is targeted for NMD.

The NMD surveillance complex consists of three proteins, Nam7, Nmd2, and Upf3. Mutations in any of the three NMD proteins cause increased translational read through and stability of mRNA containing premature stop codons (Mitchell and Tollervey. 2003, Leeds et al. 1991, Cui et al. 1995, Johansson and Jacobson. 2010). Nam7, a 109 kDa RNA helicase, associates with eRF1 and eRF3 during translation termination, and once the nascent polypeptide is released, eRF1 dissociates from the ribosome and Nmd2, or Upf3, binds the Nam7/eRF3 complex (Gonzalez et al. 2001). Nam7 binds the prematurely terminated mRNA and targets it for degradation.

Regular full-length mRNAs generally get degraded first by deadenylation of the

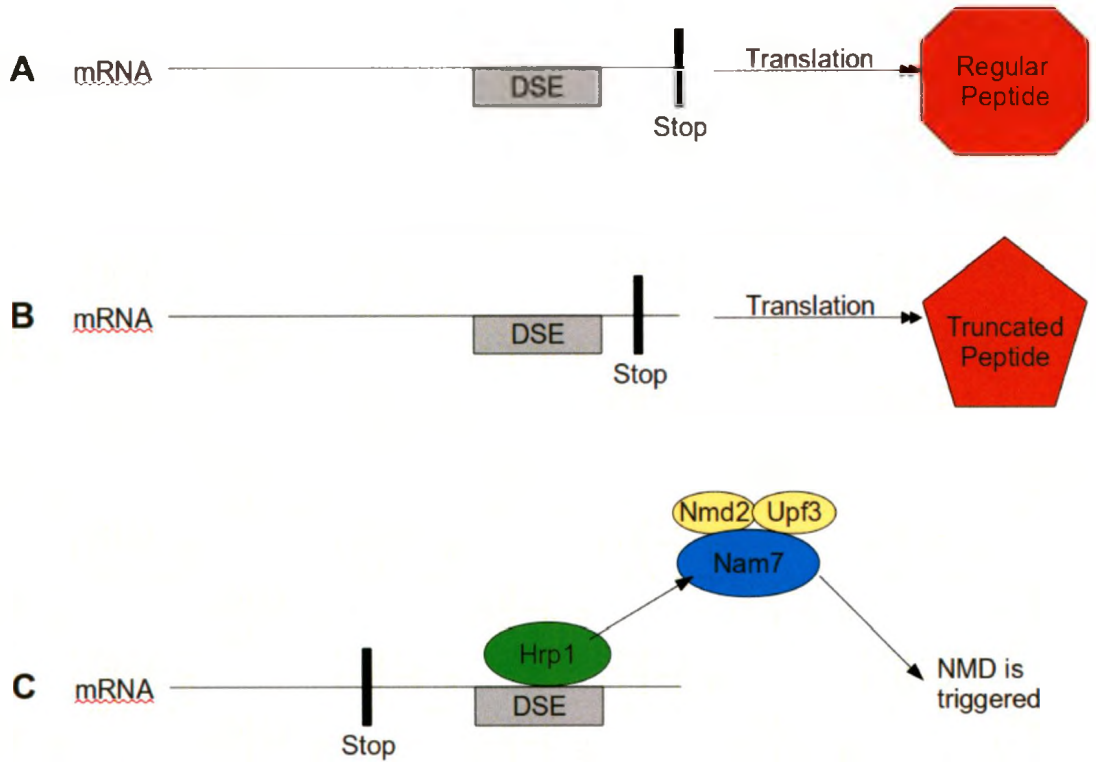

Figure 4: How Nonsense Mediated mRNA Decay (NMD) is triggered. The downstream sequence element (DSE) is a consensus sequence found towards the end of the mRNA in yeast. A) When a stop codon is located at the correct position on the mRNA, the regular peptide is properly translated. B) When a premature stop codon occurs downstream of the DSE sequence on mRNA, a truncated peptide is translated. C) When a premature stop codon occurs upstream of the DSE sequence on mRNA, the Hrp1 protein is able to bind to the DSE sequence and recruit the surveillance complex. The surveillance complex consists of Nam7, Nmd2, and Upf3, which trigger NMD.

poly(A) tail, followed by removal of the 5' methylguanosine cap and 5' to 3' degradation (Gonzalez et al. 2001). However, prematurely terminated mRNAs are degraded by two different processes in yeast. In one pathway, the 5' methylguanosine cap is removed by the Dcp1/Dcp2 complex, followed by 5' to 3' exonucleolytic decay of the mRNA by Xrn1 (Gonzalez et al. 2001, Mitchell and Tollervey. 2003). In another pathway, deadenylation first occurs by Rrp4 and the Ski complex, followed by 3' to 5' exonucleolytic decay by the Ski complex (Mitchell and Tollervey. 2003). Regardless of which pathway takes place, both the 5' and 3' NMD pathways require the surveillance complex (Nam7, Nmd2, Upf3) to first recognize these mRNAs as premature. Both pathways are inhibited in *nam7*Δ strains, or by drugs such as cyclohexamide that block translation (Gonzalez et al. 2001).

The NMD process is highly conserved in all eukaryotes (Mitchell and Tollervey. 2003). NMD is an important process for monitoring these premature stop codons so that deleterious proteins are not encoded. For example, several forms of muscular dystrophy occur due to a truncated or non-functional form of the protein dystrophin caused by a premature stop codon (Bidou et al. 2004). Therefore, further understanding of NMD is important for our understanding of diseases.

7. Concluding Remarks

Although the function of Tra1 is unknown, its presence in multiple transcriptional complexes, and as a member of the PIKK family of cellular regulators suggests it has important cellular roles. The C-terminus of Tra1 is important, as $tra1_{LA}$ results in severe growth phenotypes, transcriptional profile changes, and decreased protein levels of Tra1.

Therefore, the majority of this thesis is based on experiments performed using $tra1_{LA}$. The goal of this thesis is to examine genetic interactions of the FATC domain of Tra1 to gain more information on Tra1's cellular role. My first aim was to identify suppressor mutations of $tra1_{LA}$ that would link Tra1 to other molecules and cellular functions. A suppressor genetics approach was taken, using $tra1_{LA}$. These experiments revealed Nam7 to be a suppressor of $tra1_{LA}$, linking Tra1 to NMD. Another suppressor of a second FATC domain mutant, $tra1_{FA,}$ was revealed to be Tti2, linking Tra1 to the ASTRA complex. My second aim was to identify whether the PI3K and FATC domains of Tra1 interact with other proteins. A gel filtration experiment suggests that the C-terminus of Tra1 interacts with other molecules within the cell.

Chapter 2: Materials and Methods

1. Yeast strains

The *Saccharomyces cerevisiae* strains used in this study are listed in Table S1 in the Appendix. The *MATa* strain CY4018 contains a genomic disruption of *tra1* and is maintained by a full-length plasmid copy of nine-myc (Myc_9) tagged $tra1_{LA}$ expressed from the *DED1* promoter on a *URA3* centromeric plasmid (Hoke et al. 2010). The equivalent MATα strain CY5522 was generated by mating the *MATα* wild-type *TRA1* strain CY4413 with CY4018. After sporulation, a *MATα* spore colony was isolated that required the plasmid copy of $tra1_{LA}$ for growth. Strains carrying $tra1_{LA}$ and extragenic suppressors (hereafter defined as *es* alleles) *es2* (CY5579), *es12* (CY5580), *es35* (CY5581), *es36* (CY5582), *es37* (CY5583), *es38* (CY5584), *es39* (CY5585), *es40* (CY5586), *es41* (CY5587), *es42* (CY5588), and *es43* (CY5750) were derived from CY4018 using the selection scheme described below (Section 3). The *MATα* equivalents of CY5666, CY5758, and CY5603 (*es2* $tra1_{LA}$, *es38* $tra1_{LA}$, and *es41* $tra1_{LA}$) were generated after mating CY5579, CY5584, and CY5587 with CY5522 ($tra1_{LA}$). *MATα* spore colonies carrying the suppressor were selected based on their ability to grow at high temperature and on plates containing 4 % ethanol. Strains CY5688, CY5758, and CY5691 containing wild-type *TRA1* and *es* alleles *es2*, *es38* and *es41* respectively were generated by plasmid shuffling YCPlac33-*TRA1* (CB1617) into CY5579, CY5584, and CY5587.

Yeast strains deleted for *nam7*, *upf3* and *nmd2* were derived from the diploid consortium strains BY46214, BY44702 and BY41905 (Winzeler and Davis. 1997) respectively, by selecting kan^R colonies after sporulation. *MATa* versions of *nam7Δ*::kan^R

(CY5932), *upf3Δ*::kan^R (CY5936), and *nmd2Δ*::kan^R (CY5934) were mated with the *MATα tral$_{LA}$* strain CY5967, to yield after sporulation CY5972 (*nam7Δ*::kan^R *tral$_{LA}$*), CY5983 (*upf3Δ*::kan^R *tral$_{LA}$*), and CY5996 (*nmd2Δ*::kan^R *tral$_{LA}$*) strains. Similarly, CY5933 (MATα *nam7Δ*::kan^R) was mated with CY4400 (*MATa tral$_{FA}$*), to yield after sporulation CY5974 (*MATa nam7Δ*::kan^R *tral$_{FA}$*). CY5939 (*nam7Δ*::nat^R) was obtained by transformation and integration of a clonNAT marker (Y221) in place of KanMX. Double mutant strains of *nam7Δ*::nat^R with *ada2Δ*::kan^R (CY5979), *eaf3Δ*::kan^R (CY5980), and *eaf7Δ*::kan^R (CY5976) were obtained after mating and sporulation of the diploids of crosses of CY5939 (*nam7Δ*::nat^R) with BY4282 (*ada2Δ*::kan^R), BY7143 (*eaf3Δ*::kan^R), and BY2940 (*eaf7Δ*::kan^R). Yeast transformations of pS14 and pS17 plasmids into strain BY4742 (*TRA1* MATα) were performed to create strains CY4486 (Myc$_9$-Ngg1) and CY4455 (Myc$_9$-Tra1$_{PI3K/FATC}$). pS26 was transformed into strain BY4742 to create strain CY5766 (Myc$_9$-6HIS -Tra1$_{PI3K/FATC}$). Yeast transformations were performed using a standard lithium acetate protocol (Gietz and Woods. 2002).

CE4336 and CE4048 are *E. coli* strain ER2566 (New England Biolabs) transformed with pET21a plasmids that allow expression of a 6-HIS tagged GB domain (a kind gift of Dr. Gary Shaw) and 6-HIS tagged GB domain containing Tra1$_{PI3K/FATC}$ (encoding residues 3290 to 3744 of Tra1), respectively.

2. DNA Constructs

The DNA oligonucleotides and constructs used in this study are listed in Tables S2 and S3 respectively in the Appendix. *LacZ* reporter *PHO5* (CB1503) was previously constructed (Mutiu et al. 2007). The *TRA1* (CB1617) allele is on a *TRP1* centromeric

plasmid and was previously described (Mutiu et al. 2007).

A *PstI-SstI* fragment containing the *MET3* promoter, a Myc_9 tag, and a portion of Ngg1 (CB1997) was cloned into the YEplac195 multi-copy shuttle-vector plasmid (Gietz and Sugino. 1988) to create construct pS14 (Myc_9-Ngg1). A 1442 base pair fragment of *TRA1*, starting at nucleotide 9859 and finishing 65 nucleotides downstream of the stop codon, was generated by PCR amplification using forward primer 5759-1 and reverse primer 4479-1, incorporating *NotI* and *SstI* restriction sites on either end of the piece. This fragment was cloned into pS14 using restriction enzymes *NotI* and *SstI*, replacing the Ngg1 fragment with the PI3K and FATC domains of Tra1, to give pS17 (Myc_9-$Tra1_{PI3K/FATC}$). Oligonucleotides 1880 and 1881 encoding the 6-histidine tag, were 5' phosphorylated with polynucleotide kinase and ligated into the *NotI* site of pS17 to create construct pS26 (Myc_9-6HIS-$Tra1_{PI3K/FATC}$).

3. Isolation of $tra1_{LA}$ Extragenic Suppressors

Six CY4018 ($tra1_{LA}$) 5 mL cultures were grown in YPD (Yeast Peptone Dextrose) media to stationary phase. One milliliter of these cultures were frozen (and termed "time zero") at -70 °C. Approximately 100 million cells total were plated onto ~200 YPD plates containing 4 % ethanol. These ethanol plates were incubated at 30 °C for 2-5 days. Fast-growing colonies were selected and streaked onto YPD plates containing 4 % ethanol alongside CY4018. A $tra1_{LA}$ allele on a *TRP1* centromeric plasmid (iB150) was transformed into each suppressor strain and the *URA3 Cen* $tra1_{LA}$ allele in the strains shuffled out using 5-fluoroorotic acid. All phenotypes were then re-tested, and strains that retained their ability to grow faster than CY4018 on YPD plates containing 4 % ethanol were defined as extragenic suppressors (*es* $tra1_{LA}$). Oligonucleotide 4292-1 priming

forward at nucleotide 10687 of *TRA1*, and reverse primer 5632-1 containing a mismatch to the leucine codon at nucleotide position 11197-11198 and a match to the alanine codon, were used to confirm the presence of *tra1*-L3733A in the suppressor strains.

4. Preparation of Genomic DNA for sequencing

Yeast strains CY4018 ($tra1_{LA}$) time zero control, CY5579 (*es2* $tra1_{LA}$), CY5584 (*es38* $tra1_{LA}$), and CY5587 (*es41* $tra1_{LA}$) were grown in 15 mL YPD media to stationary phase. Cultures were transferred to sterile falcon tubes and cells collected by centrifugation for 5 minutes at 2000 xg at room temperature. Pellets were suspended in 1mL sorbitol solution (0.9 M sorbitol, 0.1 M Tris-Cl, 0.1 M EDTA, pH 8), transferred to microfuge tubes, cells pelleted by centrifugation for 4 minutes at 2000 xg , then frozen at -20 °C for 1 hour. Cells were resuspended in 0.5 mL sorbitol solution containing 100 µL of 9 mg/mL lyticase and 0.28 M beta-mercaptoethanol, incubated for 110 minutes at 35 °C with slow shaking, and for 10 minutes in a water bath at 37 °C. The resulting spheroplasts were pelleted by centrifugation at 2000 xg for 5 minutes. Spheroplasts were resuspended in 0.5 mL TE (10 mM Tris, 1 mM EDTA, pH 8) and 50 µL of 20 % SDS, and incubated for 30 minutes at 65 °C. Following the incubation, 200 µL 5 M potassium acetate was added and samples incubated for 30 minutes on ice. Samples were centrifuged at 12000 xg at room temperature, supernatants transferred to microfuge tubes, which were then filled with 95 % ethanol. Tubes were inverted, briefly centrifuged, and supernatant discarded. The resulting nucleic acid mix was suspended in 300 µL of TE (10 mM Tris, 1 mM EDTA, pH 8). Four microlitres of 50 mg/mL RNase was added to each tube and incubated for 1 hour at 37 °C. Five hundred microlitres of isopropanol was added to each tube to precipitate the DNA. DNA was spooled using a sterile needle and

suspended in 200 µL TE. A standard phenol-chloroform extraction was used to purify DNA (Chomczynski and Sacchi. 1987). Twenty microlitres of 3 M sodium acetate and 500 µL 95 % ethanol was added to each tube. Tubes were centrifuged at 13,000 xg for 3 minutes and the DNA resuspended in 50 µL sterile water. Concentrations of DNA were measured at OD_{260} using a Nanophotometer (Montreal Biotech Inc).

5. Bioinformatic analysis and Identification of $tra1_{LA}$ and $tra1_{FA}$ suppressors

Four to nine micrograms of genomic DNA from $tra1_{LA}$ time-zero control, and suppressors *es2* $tra1_{LA}$, *es38* $tra1_{LA}$, and *es41* $tra1_{LA}$ were sent to the Centre for Applied Genomics (Toronto, Ontario) for paired end sequencing using the Applied Biosystems SOLiD 4.0 next-generation platform. Genomic DNA from a $tra1_{FA}$ control plus 2 extragenic suppressors of $tra1_{FA}$ (discovered by Dr. Chris Brandl) were sequenced in the same lane. DNA library construction and barcoding was performed at the Centre. Sequencing results (colourspace data and quality files) were returned for analysis.

The *Saccharomyces cerevisiae* genome sequence was downloaded from the *Saccharomyces* Genome Database (SGD) on March 24, 2011. Custom Shell and Perl scripts were used for the sequencing analysis. The program Bowtie was used to map the colourspace reads to each chromosome of the yeast genome and obtain mapped reads in SAM (Sequence Alignment/Map) format (Langmead et al. 2009). The VCF (variant call format) from SAMtools was used to obtain a raw list of polymorphisms from the mapped reads. Reads with a Phred quality score below 20 were eliminated. A custom Perl script was written to eliminate the background polymorphisms found in wild-type samples ($tral_{LA}$ and $tral_{FA}$). The GenBank file of each yeast chromosome was downloaded from

SGD on April 12th, 2011 and a custom Perl script was used to assign gene names and descriptions to each polymorphism. Synonymous substitutions and mutations occurring in intergenic regions were manually eliminated to identify suppressors. A portion of *NAM7* was PCR amplified using forward primer 6076-1 located at nucleotide 371 of *NAM7*, and reverse primer 6076-2 located at nucleotide 622 of *NAM7*, yielding a 252 bp PCR fragment.

6. Growth Assays

Yeast strains were grown in 2 mL YPD media or minimal media containing the appropriate amino acids for plasmid maintenance. Strains were grown at 30 °C to stationary phase. Five microlitres of each strain was plated in 10-fold serial dilutions on plates containing varying concentrations of methionine (0 μM, 5 μM, 10 μM, 20 μM), rich media (YPD), or YPD containing 4 % ethanol, 1.5 μg/mL tunicamycin, 0.075 % tert-butylhydroperoxide, 1 M NaCl, low phosphate, and 2 μg/mL chloramphenicol. Plates were incubated at 30 °C unless otherwise stated.

7. β-galactosidase assays

LacZ reporter *PHO5* (CB1503) was transformed into yeast strains CY2706, CY4018, CY5579, CY5584, CY5587, CY5688, CY5690, CY5691 for β-galactosidase assays. Strains were grown to stationary phase in 2 mL synthetic dropout (SD) media lacking leucine. Cells were washed 3 times with 3 mL sterile H_2O and resuspended in 200 μL YPD depleted of phosphate. One hundred microlitres of cells were added to 3 mL of YPD depleted of phosphate. Cells were grown for 16 hours prior to performing the β-galactosidase assay (Han et al. 1988). β-galactosidase assays were performed as

previously described (Hoke et al. 2010, Mutiu et al. 2007).

8. Western blot analysis

Yeast strains CY2706 (*TRA1*), CY4018 ($tra1_{LA}$), and CY5579 ($nam7_{W\text{-}STOP}$) were grown in YPD media to OD_{600} ~2.0. Strains CY4486 (6HIS-GB) and CY4455 (Myc_9-$Tra1_{PI3K/FATC}$) were grown in both SD media lacking uracil and methionine, or lacking uracil and containing 20 µM methionine to OD_{600} ~2.0. Yeast extracts were prepared by glass bead disruption as previously described (Hoke et al. 2010). Protein concentrations were determined using Bradford protein assay reagent (BioRad) and measuring absorbance using an LKB Ultrospec Plus spectrophotometer (Pharmacia) at 595 nm. Proteins were separated using SDS-PAGE, transferred to PVDF membranes, and Western blotting procedures were performed as previously described (Mutiu et al. 2007). The following antibodies were used: monoclonal anti-c-myc (Sigma) primary antibody at a 1:4000 ratio, monoclonal anti-5HIS (QIAGEN) primary antibody at a 1:4000 ratio, and anti-mouse IgG HRP (Promega) secondary antibody at a 1:10000 ratio. Protein signals were detected using Supersignal West Pico Chemiluminiscent Substrate (Thermo Scientific).

9. TRA1 mRNA Levels

RNA was extracted from CY4434 (*TRA1*), CY5967 ($tra1_{LA}$), CY5932 ($tra1_{LA}$), and CY5972 ($tra1_{LA}\,nam7\Delta$) as previously described (Mutiu and Brandl. 2005). Concentrations of RNA were measured at OD_{260} using a Nanophotometer (Montreal Biotech Inc). The High Capacity cDNA RT Kit (Applied Biosystems) was used to obtain

cDNA from the samples following the manufacturer's instructions. Two-fold serial dilutions of 35 ng/μL of cDNA from each sample were prepared and 1 or 2 μL of each dilution was amplified by PCR. PCR reactions were carried out in 25 μL containing 25 pmol of each primer, and 1X PCR-EZ D-PCR Master Mix Kit (Bio Basic Inc.). PCR reaction conditions included a 5 minute denaturation step at 94 °C, and either 25 or 32 cycles with a 45 second denaturation step at 94 °C, a 45 second annealing step at 54 °C, and a 1 minute extension time at 72 °C. PCR products were separated by electrophoresis on a 1 % agarose gel, and visualized under UV light after staining with ethidium bromide. Forward primer 6108-1 located at nucleotide 72, and reverse primer 6108-2 located at nucleotide 524, were used to amplify a 452 bp portion of the *IDH1* gene. Forward primer 6078-1 located at nucleotide 10309, and reverse primer 4225-1 located at nucleotide 9711, were used to amplify a 599 bp portion of *TRA1*.

10. Colony PCR

A small colony from the strain of interest was incubated in a 1.5 mL microfuge tube with 10 μL of 1 μg/mL lyticase solution in TE (10 mM Tris, 1 mM EDTA, pH 8) for 30 minutes at 37 °C, followed by a 20 minute incubation at 95 °C. The tubes were centrifuged for 10 minutes at 13,000 xg and 1 μL of the supernatant was used in the PCR. PCR mixtures (total volume 25 μL) contained 1 μL of DNA (supernatant), 13.8 μL sterile H_2O, 5 μL 5X GoTaq® Flexi buffer (Promega), 2 μL 2.5 mM dNTP mix (Promega), 1.5 μL 50 mM $MgCl_2$ (Promega), 0.2 μL of GoTaq® Flexi *Taq* DNA polymerase (Promega), and 1 μL of 25 pmol of each primer. PCR reaction conditions included a 5 minute denaturation step at 94 °C, and 25 cycles with a 45 second denaturation step at 94 °C, a 45 second annealing step at 54 °C, and a 1 minute extension time at 72 °C. PCR

products were separated by electrophoresis on a 1 % agarose gel, and visualized under UV after ethidium bromide staining to check for the desired band.

11. Gel Filtration Chromatography

Yeast strain CY5766 (Myc_9-6HIS -$Tra1_{PI3K/FATC}$) was grown in SD media lacking uracil and methionine to OD_{600} ~2.0. Cells were pelleted at 4 °C and washed in PO_4 Buffer 150 (50 mM PO_4 pH 7.4, 150 mM NaCl). Yeast extract was prepared by grinding in liquid nitrogen as previously described (Hoke et al. 2010). Cell lysate of *E. coli* CE4048 (6HIS-GB) was prepared as described in the phage display section (see below), except PO_4 Buffer 150 was used. Protein concentrations were determined using the Bradford assay (BioRad). Sephadex® G-75 gel filtration beads were hydrated in TE (10 mM Tris, 1 mM EDTA, pH 8) at 90 °C, and equilibrated with PO_4 Buffer 150. Five hundred micrograms of protein was loaded onto 500 µL hydrated Sephadex® G-75 gel filtration beads in spin columns, washed four times with PO_4 Buffer 150, and fractions collected after each wash by centrifugation at 2000 xg for 1 minute. Twenty microlitres of each fraction was loaded onto two 10 % SDS-PAGE gels, transferred to PVDF membranes, and Western blotted as described above.

12. Phage Display

E. coli strains CE4336 (6HIS-GB) and CE4048 (6HIS-GB-$Tra1_{PI3K/FATC}$) were grown in 1 L LB media containing 100 µg/mL ampicillin at 37 °C to an OD_{600} ~ 0.9. Cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 hours at 15 °C. Cells were pelleted, washed with PO_4 Buffer 200 (50 mM PO_4, pH 7.4, 200 mM NaCl, 10 % glycerol), re-suspended in 30 mL PO_4 buffer 200, and stored at -70 °C

overnight. Cell suspensions were thawed on ice and protease inhibitors (1:100 BTI, 1:100 PMSF, 1:1000 leupeptin, 1:1000 pepstatin) added. Cells were homogenized using a dounce homogenizer for 5 minutes on ice, followed by lysis using a mechanical EmulsiFlex C homogenizer (Avestin). Cell lysates were collected by centrifugation for 25 minutes, 30000 xg at 4 °C. Each lysate was incubated with a 50 % IgG sepharose™ 6 Fast Flow (Amersham Biosciences) resin slurry (equilibrated in PO_4 buffer 200 with protease inhibitors) for 90 minutes rotating at 4 °C. Resins were centrifuged at 1000 xg at 4 °C for 2 minutes and the unbound fractions collected. IgG sepharose resins were incubated with 5 mg/mL BSA in TBS-T (0.5 % Tween20 in 50 mM Tris-Cl, 150 mM NaCl, pH 7.5) for 5 minutes. The resins were then equilibrated with TBS-T to prepare for phage display.

Phage display was performed using the Ph.D™-12 Phage Display Peptide Library Kit (NEB) as described by the manufacturer with the following modifications. The phage library was incubated with 6HIS-GB IgG beads for 30 minutes rotating at 4 °C and the unbound phage collected. Unbound phage were transferred directly to 6HIS-GB-Tra1$_{PI3K/FATC}$ IgG beads and incubated for 30 minutes. The 6HIS-GB-Tra1$_{PI3K/FATC}$ column was washed 12 times with TBS-T, and phage eluted by the addition of 1 mL Elution buffer (0.2 M glycine-HCl, pH 2.2, 1 mg/mL BSA). Following a 25 minute incubation at 4 °C, 150 µL of 1 M Tris-HCl, pH 9.0 was added. Eluted phage were amplified in *E. coli* strain ER2738 and precipitated using 20 % PEG/2.5 M NaCl. Phage were titered as described in the Ph.D.™ Phage Display manual (NEB) and five subsequent panning rounds were performed. Blue plaques were chosen after the final panning round and sequenced as described in the manual. Once phage were eluted from

both 6HIS-GB-Tra1$_{PI3K/FATC}$ and 6HIS-GB, IgG beads were suspended in 200 µL 1X Laemmli buffer (63 mM Tris HCl, 10 % Glycerol, 2 % SDS, 0.0025 % Bromophenol Blue, pH 6.8) and heated at 65 °C for 10 minutes to elute proteins.

Chapter 3

1. Rationale

Tra1/TRRAP are members of the PI3K-related kinase (PIKK) family of proteins, a group of molecules that function as key cellular regulators in diverse pathways. Members of this protein family contain a phosphatidylinositol-3-kinase (PI3K) domain followed by a FRAP-ATM-TRRAP C-terminus (FATC) domain at the C-terminus. The FATC domain in other members of the PIKK family is required for regulating the kinase activity of the PI3K domain (Takahashi et al. 2000, Morita et al. 2007). Interestingly, the PI3K domain of Tra1/TRRAP does not possess the kinase activity found in other regulators in the PIKK family (Saleh et al. 1998), therefore making the question of the function of the PI3K and FATC domains of Tra1/TRRAP very intriguing.

The high degree of conservation of the PI3K/FATC domains across species implies that the C-terminus of Tra1 is important for function. Indeed, mutagenesis studies have clearly shown that both the PI3K and FATC domains are important (Mutiu et al. 2007, Hoke et al. 2010). A triple alanine scanning mutation in the PI3K domain of Tra1 ($tra1_{SRR}$), results in phenotypes such as ethanol and temperature sensitivity and shortened telomeres. A phenylalanine to alanine change at position 3744 at the end of the FATC domain of Tra1 ($tra1_{FA}$) results in severe phenotypes and illustrates the importance of the extreme C-terminus of Tra1 (Hoke et al. 2010). Furthermore, addition of a terminal glycine at the end of the FATC domain results in loss of viability in yeast, illustrating that precise positioning of the C-terminus of Tra1 is important for function (Hoke et al. 2010). Although the role of the FATC domain in Tra1 has not been defined, it is clearly critical for the function of the protein.

An alignment of the FATC domains of Tra1 from various species, and from other members of the PIKK family (Hoke et al. 2010) illustrates that leucine at position 3733 is conserved throughout the entire PIKK family. When this leucine is mutated to alanine ($tra1_{LA}$), yeast are sensitive to various conditions of cellular stress, including ethanol and temperature extremes. In addition, $tra1_{LA}$ changes the cellular transcription profiles relative to the wild-type allele and results in decreased protein levels of Tra1 (Hoke et al. 2010). Consistent with this protein being critical for function, an equivalent mutation to $tra1_{LA}$ in *Drosophila* SMG1 reduced kinase activity of the adjacent PI3K domain (Morita et al. 2007). Therefore, this $tra1_{LA}$ mutation was of great interest in elucidating the function of Tra1.

Isolation of extragenic suppressor mutations of $tra1_{LA}$ provides a powerful approach to indicate intermolecular interactions, regulatory proteins, and parallel pathways that the FATC domain is involved. A genetics approach was taken to avoid any bias in determining function and because of the ease of genetics in the yeast system. Furthermore, the costs of whole-genome sequencing have decreased in the past few years, further facilitating this approach. For this chapter, suppressor genetics was used to identify and characterize extragenic mutations that suppress phenotypes caused by $tra1_{LA}$. ***My aim was to identify suppressor mutations of $tra1_{LA}$ that would link Tra1 to other molecules and cellular functions.***

2. Results

2.1 Extragenic Suppressors partially restore growth defects of *tra1$_{LA}$*

Tra1 contains a FATC domain at its C-terminus. A leucine residue at the position equivalent to 3733 of Tra1 is conserved throughout the FATC domains of the entire PIKK family, as shown in Figure 5. When this leucine is mutated to alanine (*tral$_{LA}$*), the resulting yeast strain is sensitive to ethanol and temperature extremes, shows transcriptional defects at the *PHO5* promoter, and reduced protein levels of Tra1 (Hoke et al. 2010). Therefore, *tral$_{LA}$* was used to discover extragenic suppressor mutations that may reveal clues into the functions of Tra1.

Potential extragenic suppressors were initially isolated by plating yeast strain CY4018 (*tra1$_{LA}$*) on YPD plates containing 4 % ethanol. Since CY4018 grows slowly on 4% ethanol plates, I selected for colonies that grew quickly on these plates. Forty three colonies were initially identified. Of these, 11 grew faster than CY4018 on a second screening on 4 % ethanol plates. The suppressor mutations were confirmed to be extragenic by plasmid shuffling. A *tra1$_{LA}$* allele on a *TRP1* centromeric (*CEN*) plasmid (iB150) was transformed into each suppressor strain and the *URA3 CEN tra1$_{LA}$* allele shuffled out using 5-fluoroorotic acid. All 11 strains containing extragenic suppressor mutations, retained their ability to grow faster than CY4018 on 4 % ethanol plates, with each suppressor allele tentatively referred to as *es2, es12, es35,* etcetera. Furthermore, the presence of *tra1$_{LA}$* in the suppressor strains was confirmed by colony PCR analysis. Primers 4292-1 and 5632-1 (containing a mismatch to the leucine codon and match to alanine) were used to generate a 512 bp product for *tra1* alleles that contained the L3744A mutation (Figure S1, Appendix).

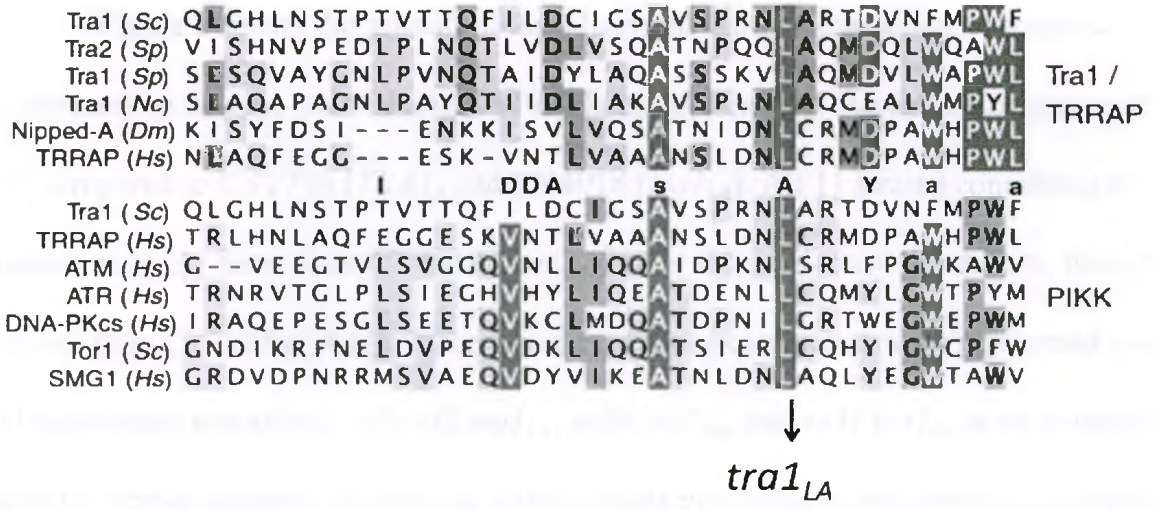

Figure 5: Tra1 FATC domain alignment showing location of *tra1*$_{LA}$ mutant (Figure from: Hoke et al. 2010). The *tra1*$_{LA}$ mutation is positioned at a leucine residue (position 3733 in *S. cerevisiae* Tra1) that is conserved throughout the entire PIKK family. Sequences of *Saccharomyces cerevisiae* Tra1, *Schizosaccharomyces pombe* Tra2, *Schizosaccharomyces cerevisiae* Tra1, *Neurospora crassa* Tra1, *Drosopihila melanogaster* Nipped-A, and *Homo sapiens* TRRAP were aligned, shown in the top panel (Hoke et al. 2010). Sequences of *Saccharomyces cerevisiae* Tra1, Tor1, *Homo sapiens* TRRAP, ATR, DNA-PKcs, and SMG1 were aligned, shown in the bottom panel (Hoke et al. 2010).

Figure 6 shows the growth of the 11 original strains containing extragenic suppressor mutations (*es* $tra1_{LA}$) on YPD plates containing 4 % ethanol, or grown at 37 °C compared to CY2706 (*TRA1*) and CY4018 ($tra1_{LA}$). All 11 strains containing *es* mutations grew faster than CY4018 ($tra1_{LA}$) under these selective conditions, though slower than CY2706 (*TRA1*). Although subsequent experiments were performed using all 11 suppressor mutations, only *es2* $tra1_{LA}$, *es38* $tra1_{LA}$, and *es41* $tra1_{LA}$ were eventually sent for whole-genome sequencing, so the results presented in this thesis are focused on these three suppressors.

Growth assays were performed using a variety of conditions to explore which of the phenotypes displayed by $tra1_{LA}$ are alleviated by the extragenic suppressors. I also wanted to determine whether suppressor mutations in the context of wild-type *TRA1* (*es2 TRA1*, CY5688; *es38 TRA1*, CY5758; *es41 TRA1*, CY5691) would display any phenotypes (Figure 7). Strains containing extragenic suppressor mutations (*es* $tra1_{LA}$) partially suppressed slow growth due to $tra1_{LA}$ on YPD plates containing 4 % ethanol, 1.5 µg/mL tunicamycin, 0.075 % tert-butylhydroperoxide, and grown at 37 °C. The ethanol, tunicamycin, and tert-butylhydroperoxide sensitivity are characteristic phenotypes of other SAGA mutants. Partial suppression was also seen on YPD containing low phosphate, which is an indicator of defects in inducing *PHO5* genes. No suppression was seen on YPD containing 2 µg/mL chloramphenicol. Interestingly, strains containing *es* $tra1_{LA}$ grew faster than CY2706 (*TRA1*), CY4018 ($tra1_{LA}$), and strains containing the *es* suppressor mutations in the context of *TRA1* on 1 M NaCl. In the context of *TRA1* the *es* suppressor mutations resulted in similar or slower growth as compared to the wild-type in the conditions assayed. CY5691 (*es41 TRA1*) grew slower than CY2706, CY5584 (*es2*

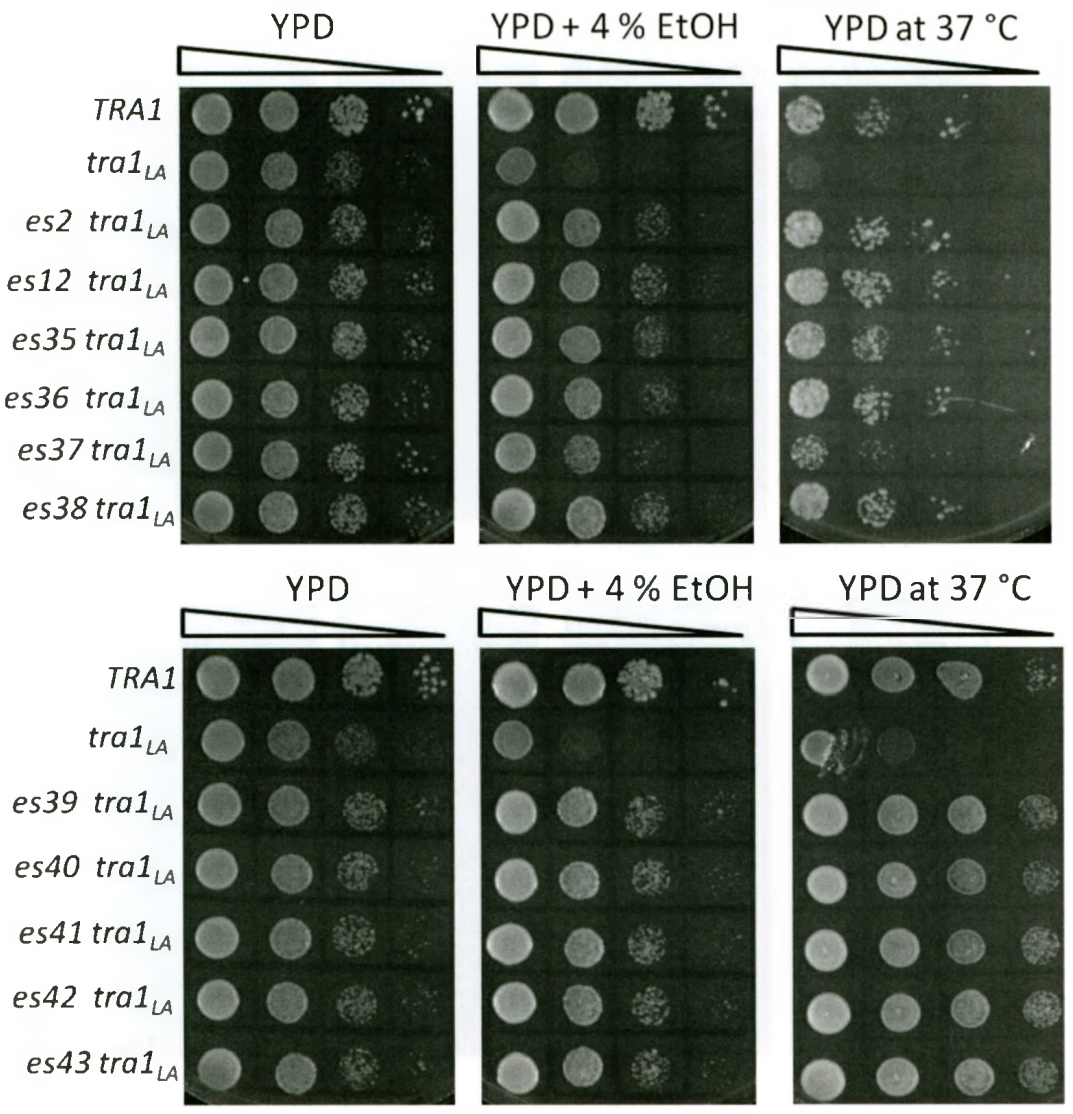

Figure 6: Original Extragenic Suppressors (*es*) of *tra1$_{LA}$*. Growth of CY2706 (*TRA1*), CY4018 (*tra1$_{LA}$*), and strains containing extragenic suppressor mutations (*es*) and *CEN* plasmids expressing *tra1$_{LA}$* were compared. Strains were grown to stationary phase and plated in 10-fold serial dilutions on rich media (YPD), and YPD containing 4 % ethanol or grown at 37 °C. Plates were incubated at 30 °C unless otherwise stated.

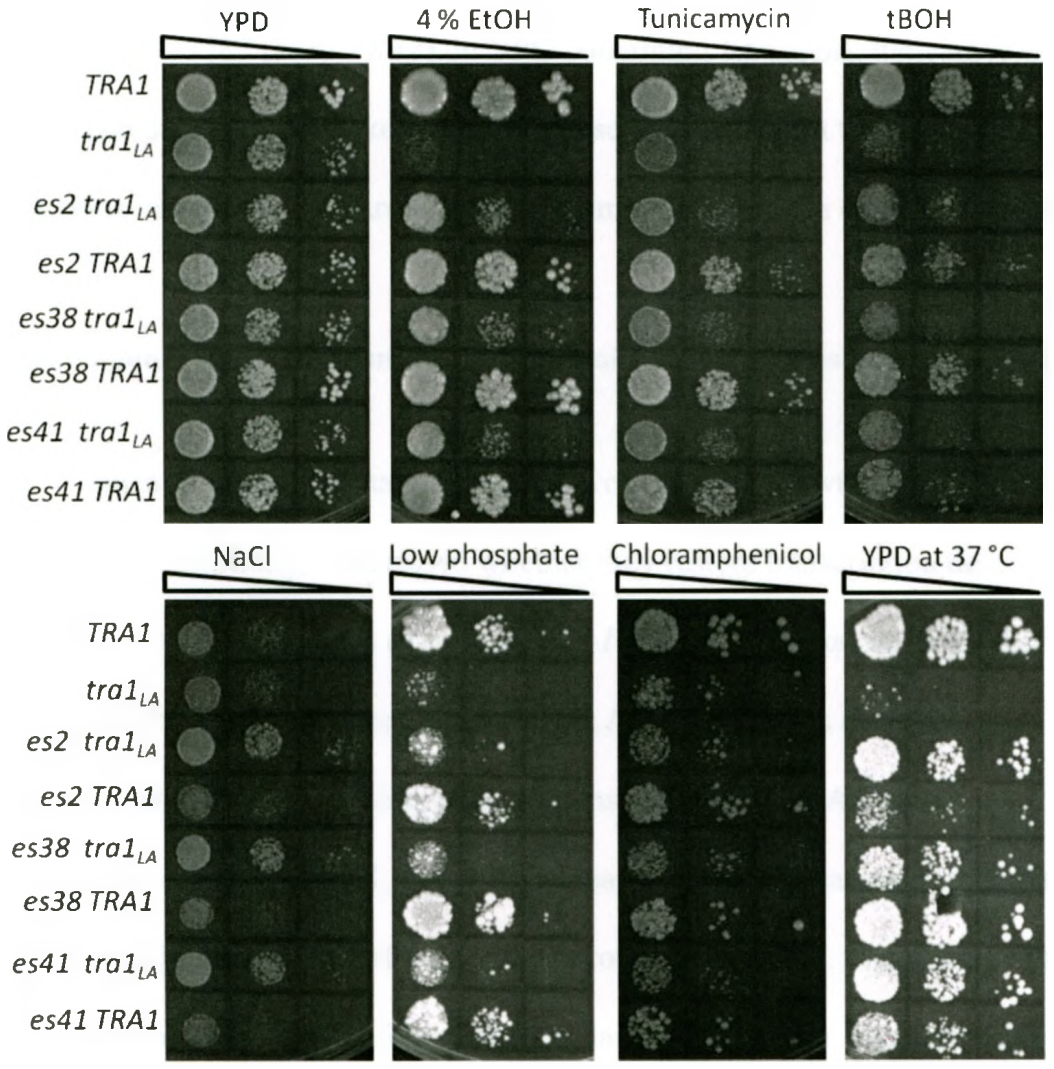

Figure 7: Three Extragenic Suppressors (*es*) of *tra1$_{LA}$*. Growth of CY2706 (*TRA1*), CY4018 (*tra1$_{LA}$*), three strains with extragenic suppressor mutations containing *CEN* plasmids expressing *tra1$_{LA}$* CY5579 (*es2 tra1$_{LA}$*), CY5584 (*es38 tra1$_{LA}$*), CY5587 (*es41 tra1$_{LA}$*), and three strains with extragenic suppressor mutations containing *CEN* plasmids expressing *TRA1* CY5688 (*es2 TRA1*), CY5690 (*es38 TRA1*), and CY5691 (*es41 TRA1*) were compared. Strains were grown to stationary phase and plated in 10-fold serial dilutions on rich media (YPD), or YPD containing 4 % ethanol, 1.5 µg/mL tunicamycin, 0.075 % tert-butylhydroperoxide (tBOH), 1 M NaCl, low phosphate, 2 µg/mL chloramphenicol, and YPD at 37 °C.

TRA1) and CY5587 (*es38 TRA1*) on YPD containing 1.5 µg/mL tunicamycin or 0.075 % tert-butylhydroperoxide. Taken together, these results suggest that the extragenic suppressor mutations (*es*) partially restore some of the growth defects observed by $tra1_{LA}$.

2.2 Extragenic Suppressor mutations partially restore transcriptional defects of $tra1_{LA}$

Since suppressor mutations partially restore slow growth resulting from $tra1_{LA}$ under stress conditions, I addressed whether they restore transcription defects associated with $tra1_{LA}$. $tra1_{LA}$ decreases expression of a *PHO5-LacZ* reporter (Hoke et al. 2010). Expression from the *PHO5* promoter using a *LacZ* fusion was determined in CY2706 (*TRA1*), CY4018 ($tra1_{LA}$), the *es* $tra1_{LA}$ strains, and the *es* TRA1 strains.

As shown in Figure 8, the *es alleles* partially restore transcriptional activation of *PHO5-LacZ* in the context of $tra1_{LA}$ by approximately 2-fold. The effects of the suppressors on transcription followed a similar pattern to their suppression of $tra1_{LA}$'s growth defects on low phosphate plates. Strains containing suppressor mutations grew faster than CY4018 ($tra1_{LA}$) on low phosphate YPD plates, but not as fast as CY2706 (*TRA1*). The ability of the suppressor mutations to suppress the transcriptional defects of $tra1_{LA}$ is thus partial. Interestingly, the strains with the extragenic suppressors in a wild-type *TRA1* background (*es TRA1*) showed a 50% decrease in transcription (compared to the wild-type strain). The transcriptional activity of $tra1_{LA}$ is about 25 % of *es2 TRA1*, *es38 TRA1*, and *es41 TRA1*, whereas the transcriptional activity of $tra1_{LA}$ extragenic suppressors (*es* $tra1_{LA}$) is about 50 % of *es2 TRA1*, *es38 TRA1*, and *es41 TRA1*.

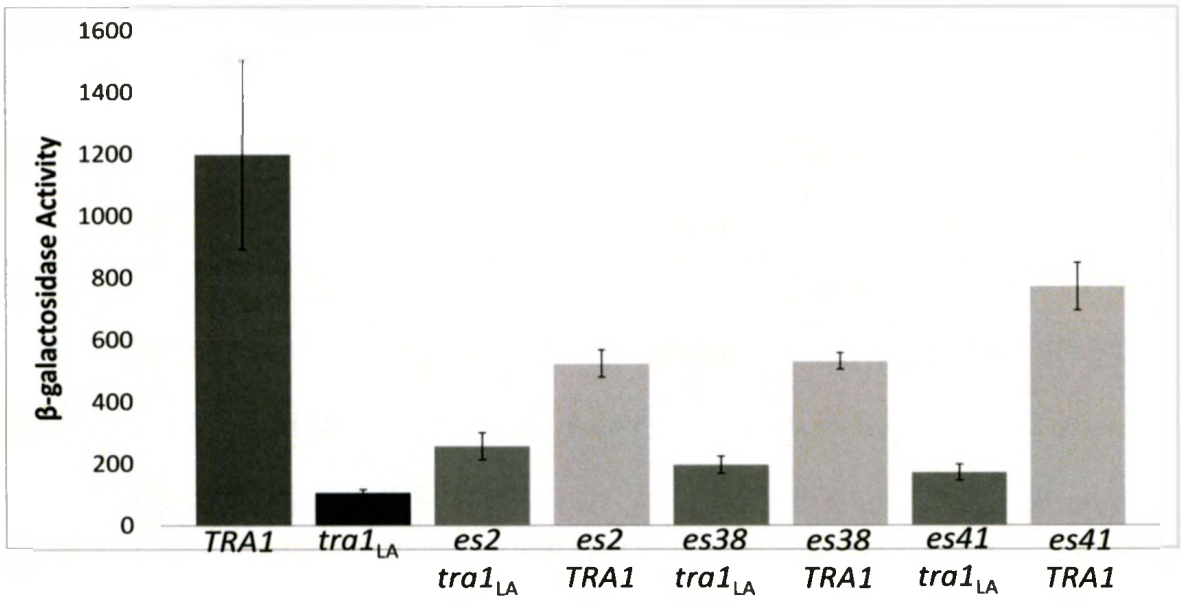

Figure 8: *ß*-galactosidase Assay. *ß*-galactosidase activity of wild-type Tra1 (*TRA1*), leucine 3733 to alanine mutant ($tra1_{LA}$), three extragenic suppressor mutations containing centromeric plasmids expressing $tra1_{LA}$ (*es2* $tra1_{LA}$, *es38* $tra1_{LA}$, *es41* $tra1_{LA}$), and the three extragenic suppressors containing centromeric plasmids expressing *TRA1* (*es2 TRA1*, *es38 TRA1*, *es41 TRA1*) were compared under the *PHO5* promoter. *ß*-galactosidase activity was determined after growth in low phosphate media for 16 hours at 30 °C. Samples were performed in triplicate, and the standard deviation is shown.

2.3 Extragenic suppressors are recessive and belong to three complementation groups

Classical genetic approaches were used to determine if the suppressor mutations were single genes, acting dominantly or recessively, and to determine the number of complementation groups. *MATa* haploid strains CY5579 (*es2 tra1$_{LA}$*), CY5584 (*es38 tra1$_{LA}$*), and CY5587 (*es41 tra1$_{LA}$*) were mated with *MATα* haploid CY5522 (*tra1$_{LA}$*) to produce diploid strains homozygous for *tra1$_{LA}$* and heterozygous for each suppressor. In this manner, I could determine whether suppressors were dominant or recessive. Figure 9 shows that the diploid strains, heterozygous for each of the suppressor mutations, grow slowly on 4 % ethanol and at 37 °C. Therefore, the suppressor mutations were recessive, indicating that a loss of function allele likely caused the suppression.

Since suppressors were recessive, complementation groups could be determined by mating the suppressor strains. Figure 10 shows the complementation group determination for *es2 tra1$_{LA}$*. *MATa* haploid strain CY5579 (*es2 tra1$_{LA}$*) was mated with the *MATα* version (CY5666) to produce a diploid strain homozygous for *es2 tra1$_{LA}$*. As expected for the *es2* suppressor mutation, the diploid homozygous for *es2* grew faster than the diploid heterozygous for *es2* on a YPD plate containing 4 % ethanol. Similarly, *MATα* CY5666 (*es2 tra1$_{LA}$*) was mated with *MATa* CY5584 (*es38 tra1$_{LA}$*) and CY5587 (*es41 tra1$_{LA}$*) to check for complementation. Both the *es38/es2* and *es41/es2* diploids grew slowly on a YPD plate containing 4 % ethanol (Figure 10), demonstrating that *es2* is in a separate complementation group from *es38* and *es41*. Similar matings were performed for all 11 haploid strains containing *tra1$_{LA}$* extragenic suppressor mutations and three complementation groups were found (results not shown). One complementation

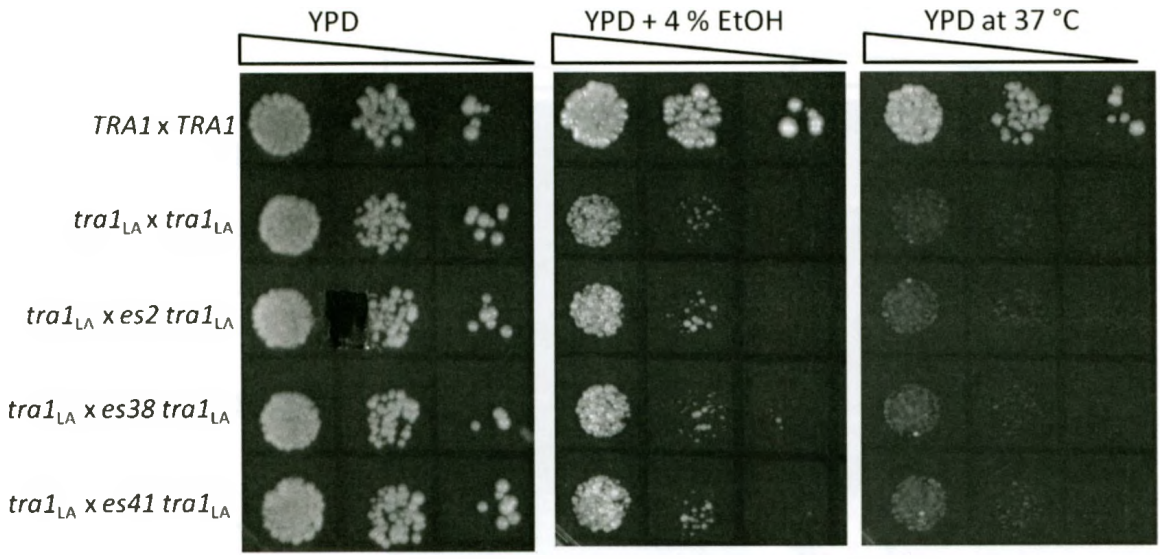

Figure 9: Genetics test of extragenic *tra1$_{LA}$* suppressors. Growth of wild-type diploid (*TRA1* x *TRA1*), leucine 3733 to alanine diploid (*tra1$_{LA}$* x *tra1$_{LA}$*), and three diploid extragenic suppressors (homozygous for *tra1$_{LA}$* and heterozygous for each suppressor *es2*, *es38*, and *es41*) were compared. Strains were grown to stationary phase and plated in 10-fold serial dilutions on rich media (YPD), and YPD containing 4 % ethanol or grown at 37 °C. Plates were incubated at 30 °C unless otherwise stated.

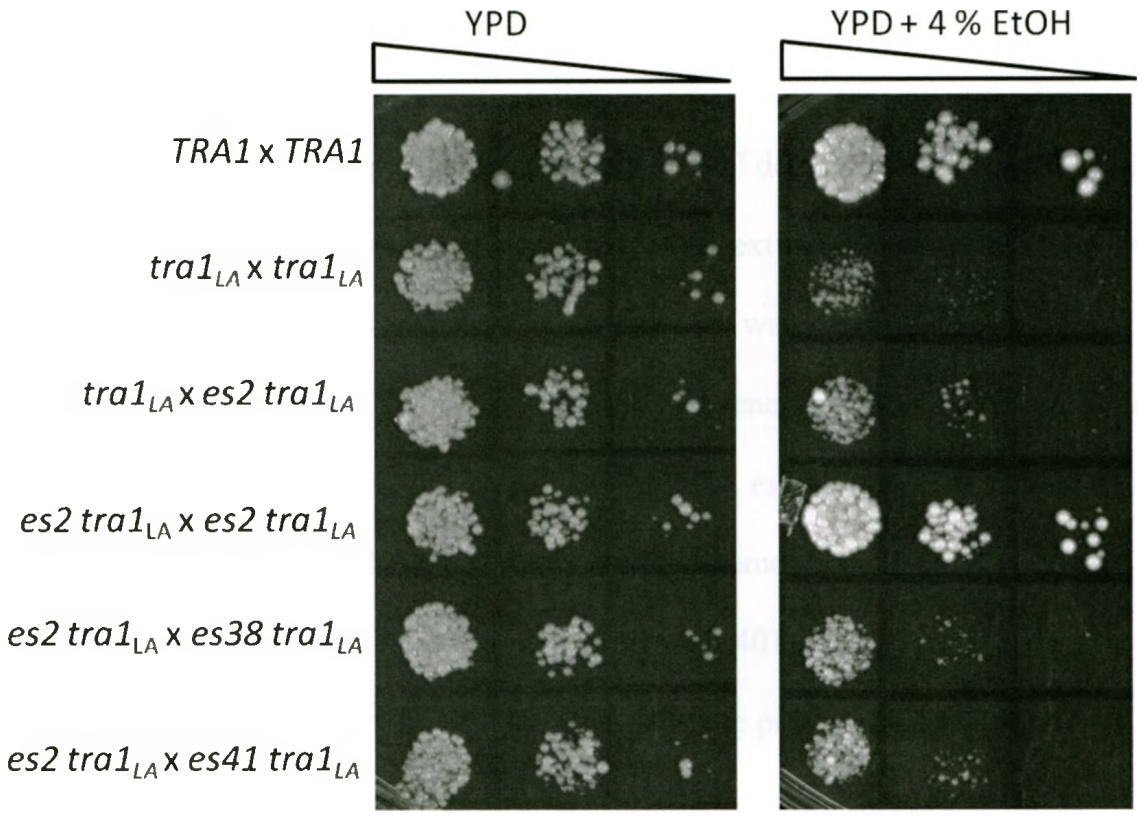

Figure 10: Complementation group of *es2* suppressor. Growth of wild-type diploid (*TRA1* x *TRA1*), leucine 3733 to alanine diploid ($tra1_{LA}$ x $tra1_{LA}$), heterozygous suppressor (*es2* $tra1_{LA}$ x $tra1_{LA}$), and diploid extragenic suppressors (homozygous for $tra1_{LA}$ and heterozygous for two suppressors) were compared. Strains were grown to stationary phase and plated in 10-fold serial dilutions on rich media (YPD), or YPD containing 4 % ethanol. Plates were incubated at 30 °C.

group consisted of *es2*, a second of *es35*, and a third group comprising the remainder of the suppressors.

Since *es2* partially suppressed the transcriptional defects of $tra1_{LA}$, and was the only one in its complementation group, genomic DNA extracted from the CY5579 strain was sent for whole-genome sequencing. Although *es35* was also in its own complementation group, its DNA was not sent for sequencing because it did not restore transcription at the *PHO5* promoter (results not shown). *es38* and *es41* were in the same complementation group. Both were sent for whole-genome sequencing to increase the chances of identifying the mutation. DNA from the CY4018 ($tra1_{LA}$) time zero control was also sent for sequencing to identify mutations in the parent strain background.

2.4 Analysis of whole-genome sequencing data and identification of suppressors

A total of 15 samples were sent to the Centre for Applied Genomics for paired end sequencing using the Applied Biosystems SOLiD 4.0 next-generation platform. DNA from the $tra1_{LA}$ time zero control was sent along with *es2*, *es38*, and *es41*. As mentioned in the introduction, a mutation at the end of the FATC domain of Tra1 ($tra1_{FA}$) also results in growth phenotypes and transcriptional defects (Hoke et al. 2010). Therefore, a control strain containing $tra1_{FA}$ plus 2 extragenic suppressors of $tra1_{FA}$ (isolated by Dr. Chris Brandl) were sequenced in the same lane. Seven additional samples came from Dr. Ivan Sadowski (University of British Columbia) and one from Dr. Greg Gloor. Figure 11 outlines the method used for analyzing the whole-genome sequencing data. Sequencing results (colourspace data and quality files) were returned, and the *Saccharomyces cerevisiae* genome was downloaded from SGD. There were approximately 50 million sequencing reads for each sample. The program Bowtie was used to map the colourspace

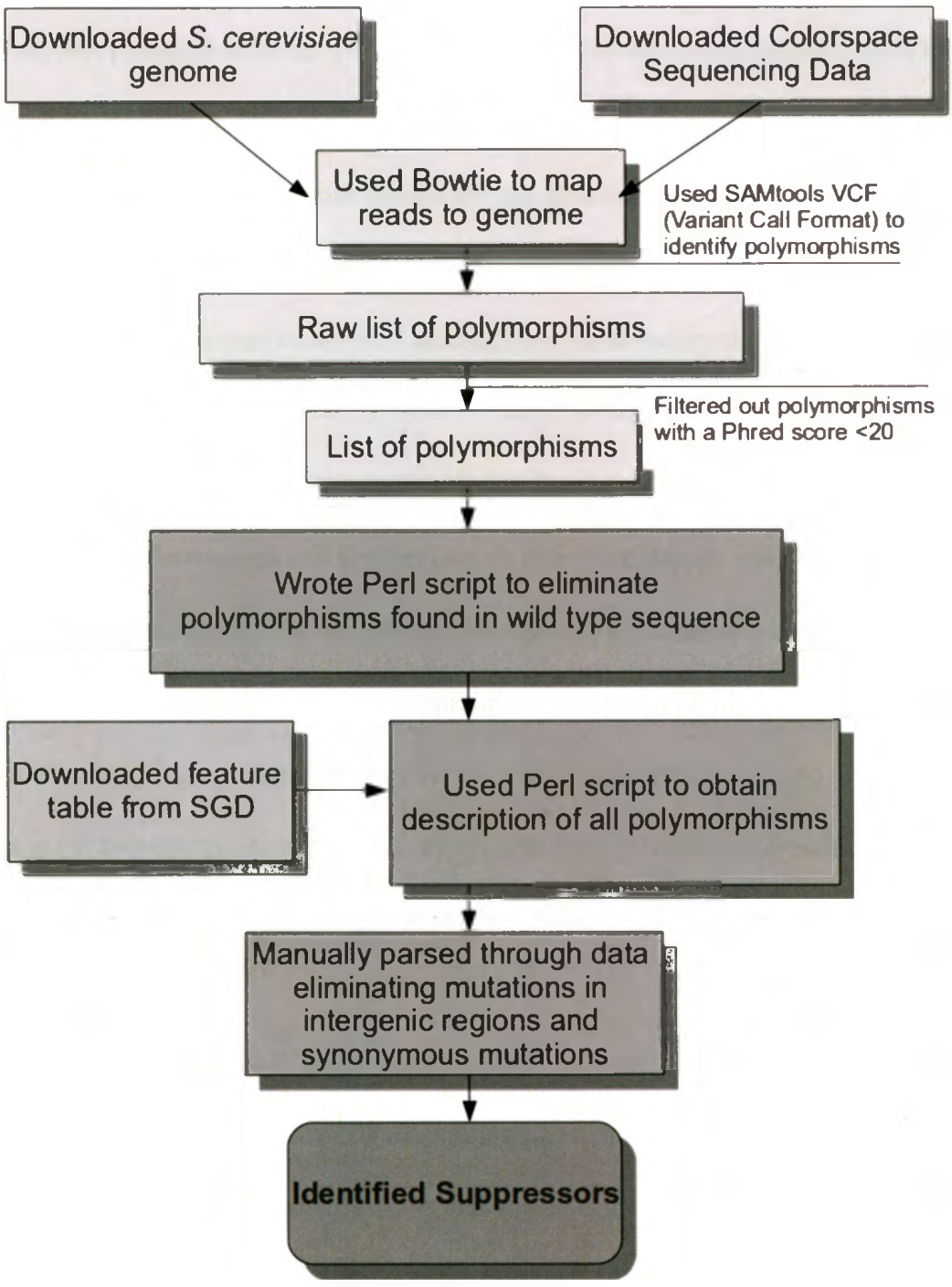

Figure 11: Identification of Extragenic Suppressor mutations using Bioinformatics.

This flow-chart describes the analysis of ABI SOLiD 4.0 colourspace sequencing data to identify $tral_{LA}$ and $tral_{FA}$ extragenic suppressor mutations.

sequencing reads to each chromosome of the yeast genome, and obtain mapped reads in SAM format (Langmead et al. 2009). Bowtie mapped approximately 60 % of the sequence reads from each sample to the yeast genome. Forward and reverse reads were mapped separately because more coverage was obtained in this manner, as opposed to mapping them as paired-ends. The "v - 3" alignment mode from Bowtie was used, allowing reads with a maximum of 3 mismatches to be mapped to the genome (Figure S2).

Each mapped read was output in SAM format. A raw list of polymorphisms was compiled using the variant call format (VCF) from SAMtools. Polymorphisms with a Phred score greater than 20 were filtered into a new file and further analyzed. A Phred score of 20 was chosen because this insures that the polymorphism occurs 99 % of the time in the sequences. A custom Perl script was written (Figure S3) to eliminate the background polymorphisms found in the parent strains ($tral_{LA}$ and $tral_{FA}$) from suppressors. Only polymorphisms that were unique to a particular complementation group were examined.

The GenBank file of each yeast chromosome was downloaded from SGD and a custom Perl script used to assign gene names and descriptions to each polymorphism (Figure S4). Synonymous substitutions and mutations occurring in intergenic regions were eliminated from the analysis. The *es2* mutation that suppressed $tral_{LA}$ was a UGG (encoding tryptophan) to UGA (opal) mutation in the *NAM7* gene at codon 165 ($nam7_{W-UM}$). Nam7 is a protein involved in mRNA decay. The mutation that suppressed $tral_{FA}$ was a T to C transition in codon 328 of *TTI2*, converting the codon from TTT (phenylalanine) to TCT (serine). Tti2 is a component of the ASTRA complex. There was

only one candidate suppressor for the *es38* and *es41* complementation group and this was a G to A transition in codon 39 of *SRB6*, converting the codon from GGT (glycine) to GAT (aspartic acid). *SRB* components are part of Mediator and they are needed for transcription of most genes *in vivo* (Martinez. 2002). After sequencing the *SRB6* gene however, I found that it was in fact a background mutation. Therefore, no suppressors were identified for the *es38* and *es41* complementation group.

To confirm that the $nam7_{W\text{-}UM}$ allele segregated with the suppression phenotype, CY5579 (*es2* $tra1_{LA}$) was mated with CY4018 ($tra1_{LA}$), the resulting diploid sporulated, and spore colonies tested for growth on YPD plates containing 4 % ethanol. DNA was extracted from four colonies exhibiting slow growth on 4 % ethanol (believed to be *NAM7* $tra1_{LA}$) and four colonies exhibiting faster growth than CY4018 ($tra1_{LA}$) on 4 % ethanol (believed to be suppressor $nam7_{W\text{-}UM}$). A portion of *NAM7* was PCR amplified and sequenced. All strains exhibiting slow growth on 4 % ethanol contained wild-type *NAM7*; all strains exhibiting faster growth than CY4018 on 4 % ethanol contained $nam7_{W\text{-}UM}$. Therefore, the *es2* suppressor mutation was confirmed to be $nam7_{W\text{-}UM}$.

2.5 Suppressor $nam7_{W\text{-}UM}$ helps restore $tra1_{LA}$ protein levels

Since $tra1_{LA}$ decreases $tra1_{LA}$ protein levels (Hoke et al. 2010), it is possible that $nam7_{W\text{-}UM}$ acts by increasing $tra1_{LA}$ protein levels. Western blotting was used to determine the amount of $tra1_{LA}$ in yeast extracts from CY2706 (*TRA1*), CY4018 ($tra1_{LA}$), and suppressor CY5579 ($nam7_{W\text{-}UM}$). Tra1 was N-terminally Myc_9-tagged in all samples, and proteins separated using SDS-PAGE (Figure 12). A band corresponding to ~433 kDa (Tra1) was present in all samples. $tra1_{LA}$ showed reduced protein levels of $tra1_{LA}$ as compared to *TRA1*. The extragenic suppressor $nam7_{W\text{-}UM}$ increased the amount of $tra1_{LA}$.

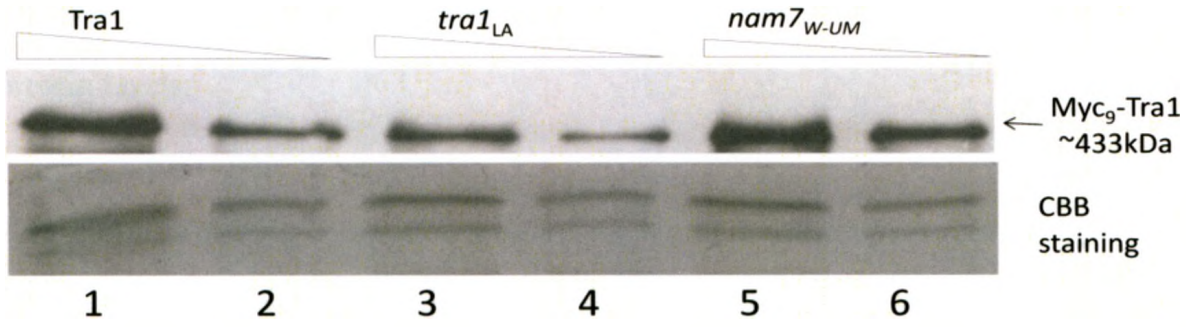

Figure 12: Tra1 protein levels in the presence of *nam7*$_{W\text{-}UM}$. Cell lysates were prepared by bead lysis from CY2706 (Myc$_9$-Tra1), CY4018 (Myc$_9$-*tra1*$_{LA}$), and CY5579 (Myc$_9$-*tra1*$_{LA}$, *nam7*$_{W\text{-}UM}$). Sixty micrograms (lanes 1, 3, and 5) and 30 µg (lanes 2, 4, and 6) of protein was separated by SDS-PAGE on a 5 % gel, transferred to a PVDF membrane, and probed with an α-myc antibody. The bottom half of the gel was stained with Coomassie Brilliant Blue (CBB).

Therefore, the *nam7*$_{W\text{-}UM}$ suppression of *tra1*$_{LA}$ may occur through restoration of *tra1*$_{LA}$ protein levels.

2.6 Deletions of NMD components *nam7*, *upf3*, and *nmd2* suppress *tra1*$_{LA}$

Since the *tra1*$_{LA}$ suppressor was a tryptophan (W) to opal mutation in the *NAM7* gene (*nam7*$_{W\text{-}UM}$), it was possible that *nam7Δ* would also suppress defects exhibited by *tra1*$_{LA}$. Suppression by *nam7Δ* would be consistent with the suppressor being recessive and a loss of function mutation. Furthermore, since Nam7, Upf3, and Nmd2, are components of the NMD machinery, I tested whether *upf3Δ* and *nmd2Δ* would also result in suppression of *tra1*$_{LA}$. Growth assays were performed on YPD plates containing 6 % ethanol or grown at 37 °C to test whether *nam7Δ*, *upf3Δ,* and *nmd2Δ*, would suppress the growth defects seen by *tra1*$_{LA}$. Six percent ethanol was used (rather than 4 %) because the consortium strain background is less sensitive to ethanol. Figure 13 shows that CY5972 (*tra1*$_{LA}$/*nam7Δ*), CY5983 (*tra1*$_{LA}$/*upf3Δ*), and CY5996 (*tra1*$_{LA}$/*nmd2Δ*) grow faster than CY5967 (*tra1*$_{LA}$) on 6 % ethanol, though equivalent to CY5967 on YPD at 37 °C. Therefore *nam7Δ, upf3Δ*, and *nmd2Δ* suppress the ethanol, but not the temperature, sensitivity phenotypes associated with *tra1*$_{LA}$.

2.7 *TRA1* mRNA levels appear to be unchanged in *tra1*$_{LA}$, *nam7Δ*, and *tra1*$_{LA}$ *nam7Δ*

Nam7 is part of the surveillance complex involved in nonsense mediated mRNA decay (NMD). Since the *tra1*$_{LA}$ suppressor *nam7Δ* helps recover *tra1*$_{LA}$ protein levels, perhaps *nam7Δ* stabilizes *TRA1* mRNA. RNA was extracted from CY4434 (*TRA1*), CY5967 (*tra1*$_{LA}$), CY5932 (*nam7Δ*), and CY5972 (*tra1*$_{LA}$ *nam7Δ*) strains to investigate whether there were differences in expression. Two-fold serial dilutions of 35 ng/μL of

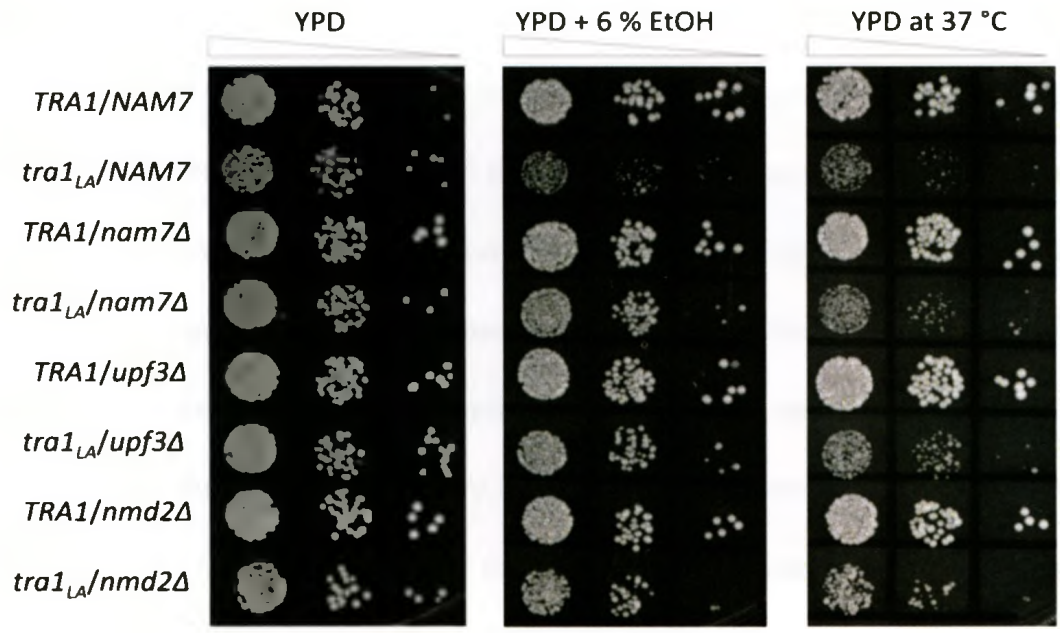

Figure 13: Growth assays testing suppression of *tra1$_{LA}$* by *nam7Δ*, *upf3Δ*, and *nmd2Δ*. Growth of CY4434 (*TRA1/NAM7*), CY5967 (*tra1$_{LA}$/NAM7*), CY5932 (*TRA1/nam7Δ*), CY5972 (*tra1$_{LA}$/nam7Δ*), CY5936 (*TRA1/upf3Δ*), CY5983 (*tra1$_{LA}$/upf3Δ*), CY5935 (*TRA1/nmd2Δ*), and CY5996 (*tra1$_{LA}$/nmd2Δ*) haploids were compared. Strains were grown in YPD media to stationary phase, and 10-fold serial dilutions were plated onto rich media (YPD), YPD containing 6 % ethanol, or YPD grown at 37 °C. Strains were grown at 30 °C unless otherwise stated.

cDNA from each sample were prepared and 1 μL (*IDH1*) or 2 μL (*TRA1*) of each dilution was amplified by PCR. *IDH1* was used as a control because previous gene expression studies (Hoke et al. 2010) indicated that *IDH1* mRNA was unchanged by the $tra1_{LA}$ mutation. Figure 14 qualitatively shows that the relative levels of mRNA between *TRA1* and *IDH1* are unchanged. Unfortunately the intensity of the bands was not sufficient to allow quantitation. Therefore, the ratios of *TRA1* mRNA and *IDH1* mRNA were qualitatively compared. The relative band intensities observed in the *TRA1* cDNA amplifications (Figure 14A) mirror the band intensities seen in the *IDH1* control amplifications (Figure 14B). Figure 14C shows a PCR water control, a negative control (no RNA), a positive control (amplification of *TRA1*), and the no reverse transcriptase controls for each sample. Bands were absent in the no RT controls, indicating that the RNA samples were free of genomic DNA. Therefore, these initial results suggest that *TRA1* mRNA levels are comparable between *TRA1*, $tra1_{LA}$, *nam7Δ*, and $tra1_{LA}$ *nam7Δ*. However further experiments should be performed to quantify these initial results.

2.8 *nam7Δ* suppressed $tra1_{FA}$ but not deletions of SAGA or NuA4 components

I tested whether *nam7Δ* would suppress another FATC domain mutation, $tra1_{FA}$, and deletions of the SAGA component *ada2* or NuA4 components *eaf3* and *eaf7*. Growth assays were performed on YPD plates containing 6 % ethanol, 6 % ethanol grown at 35 °C, or YPD grown at 37 °C. CY5974 (*nam7Δ*/$tra1_{FA}$) grows faster than CY4400 ($tra1_{FA}$) on YPD plates containing 6 % ethanol and grown at 37 °C (Figure 15). Therefore *nam7Δ* is a suppressor of $tra1_{FA}$. As shown in Figure 16, CY5979 (*nam7Δ*/*ada2Δ*) grows at the same rate as BY4282 (*ada2Δ*) on YPD plates containing 6 % ethanol and thus *nam7Δ* is

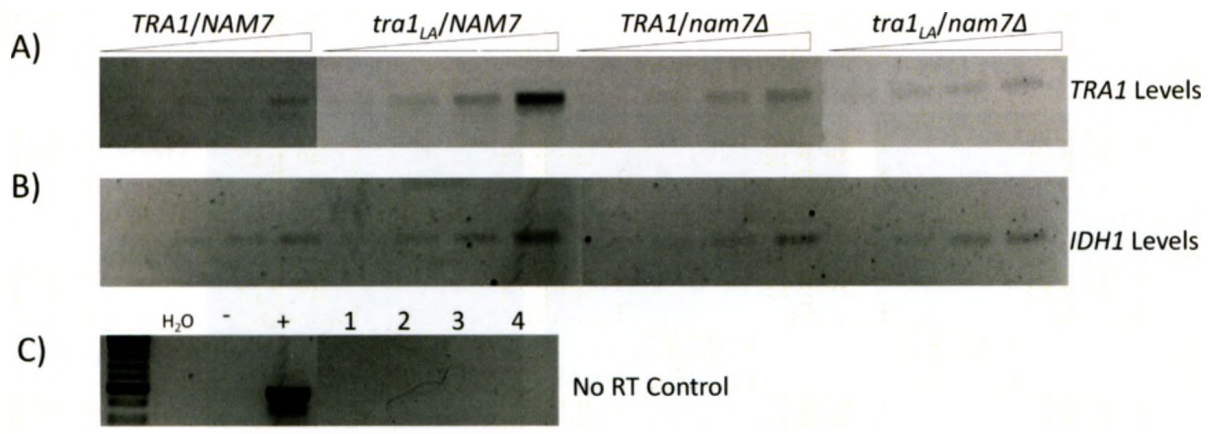

Figure 14: mRNA Levels of *TRA1* and *IDH1*. RNA was extracted from CY4434 (*TRA1*), CY5967 ($tra1_{LA}$), CY5932 (*nam7Δ*), and CY5972 ($tra1_{LA}$ *nam7Δ*) strains. The High Capacity cDNA RT Kit was used to obtain cDNA from the samples. A no RNA control, and no reverse transcriptase controls were also used for each sample. Two-fold serial dilutions of 35 ng/µL of cDNA from each sample were prepared and amplified by PCR. PCR products were separated by electrophoresis on a 1 % agarose gel, and visualized under UV light after staining with ethidium bromide. A) *TRA1* cDNA levels for all 4 samples, B) *IDH1* cDNA levels for all 4 samples, and C) PCR water control (H_2O), no RNA control (-), a positive genomic DNA control (+), and no reverse transcriptase controls for *TRA1* (1), $tra1_{LA}$ (2), *nam7Δ* (3), and $tra1_{LA}$ *nam7Δ* (4).

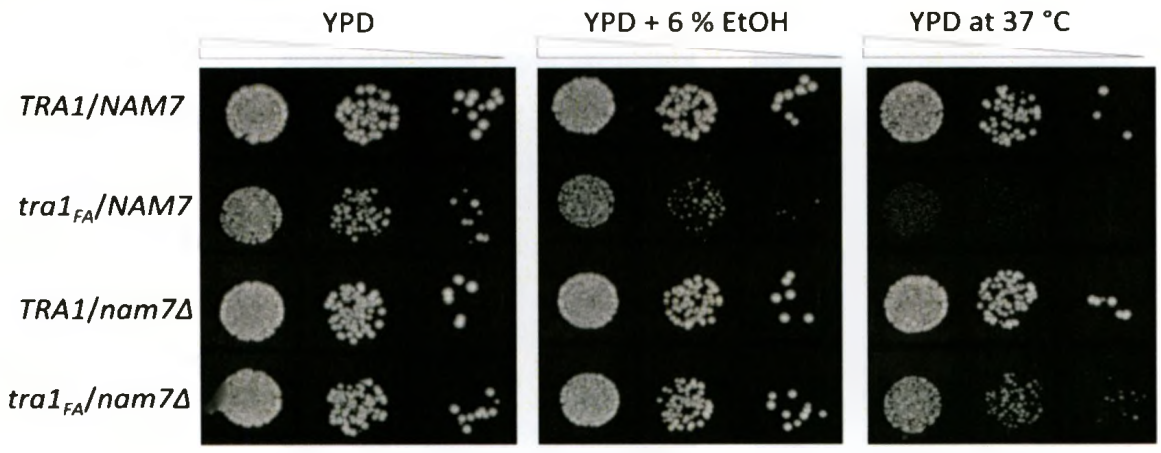

Figure 15: Growth assays testing allele specificity of *nam7Δ* with $tra1_{FA}$. Growth of CY4434 (*TRA1/ NAM7*), CY4400 ($tra1_{FA}$/*NAM7*), CY5932 (*TRA1*/*nam7Δ*), and CY5974 ($tra1_{FA}$/*nam7Δ*) haploids were compared. Strains were grown in YPD media to stationary phase, and 10-fold serial dilutions were plated onto rich media (YPD), YPD containing 6 % ethanol, or YPD grown at 37 °C. Plates were incubated at 30 °C unless otherwise stated.

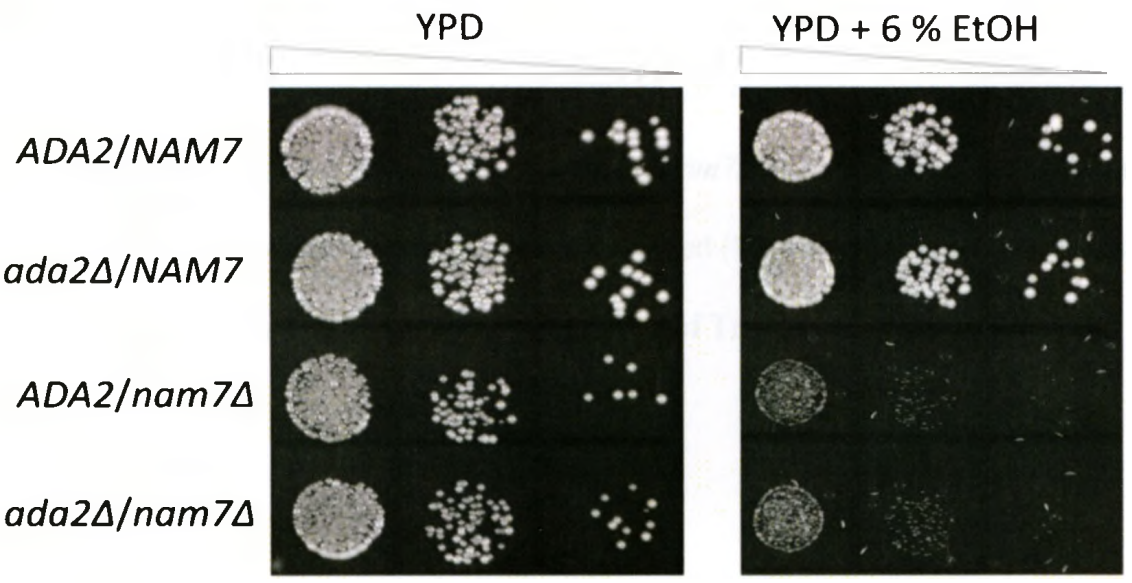

Figure 16: Growth assays testing allele specificity of *nam7Δ* with *ada2Δ*. Growth of CY4434 (*ADA2/NAM7*), BY4282 (*ada2Δ/NAM7*), CY5932 (*ADA2/nam7Δ*), and CY5979 (*ada2Δ/nam7Δ*) haploids were compared. Strains were grown in YPD media to stationary phase, and 10-fold serial dilutions were plated onto rich media (YPD), and YPD containing 6 % ethanol. Plates were incubated at 30 °C unless otherwise stated.

not a suppressor of *ada2Δ*. Interestingly, CY5980 (*nam7Δ/eaf3Δ*) and CY5976 (*nam7Δ/eaf7Δ*) grow slower than BY7143 (*eaf3Δ*) and BY2940 (*eaf7Δ*) on YPD plates containing 6 % ethanol grown at 35 °C indicating *nam7Δ* does not suppress *eaf3Δ* or *eaf7Δ*, and in fact a synthetic slow growth is observed (Figure 17). Taken together, these results suggest that *nam7Δ* is a specific suppressor of Tra1.

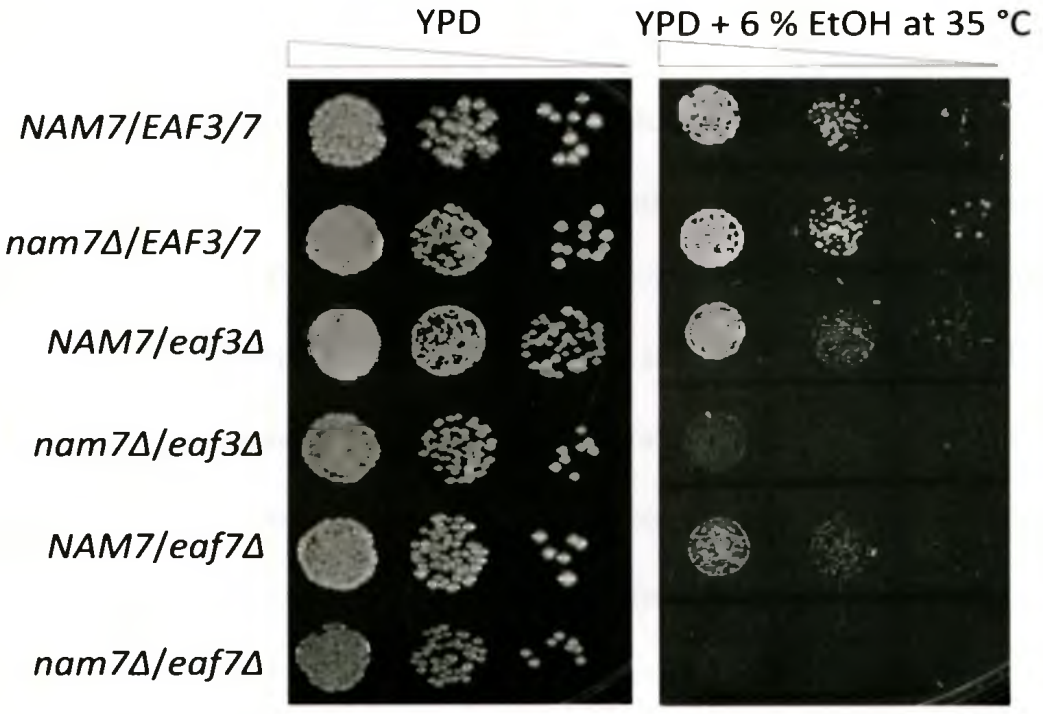

Figure 17: Growth assays testing allele specificity of *nam7Δ* with *eaf3Δ* and *eaf7Δ*. Growth of CY4434 (*NAM7/EAF3/7*), CY5967 (*nam7Δ/EAF3/7*), BY7143 (*NAM7/eaf3Δ*), CY5980 (*nam7Δ /eaf3Δ*), BY2940 (*NAM7/eaf7Δ*), and CY5976 (*nam7Δ/eaf7Δ*) haploids were compared. Strains were grown in YPD media to stationary phase, and 10-fold serial dilutions were plated onto rich media (YPD) grown at 30 °C, and YPD containing 6 % ethanol grown at 35 °C.

Chapter 4

1. Rationale

Tra1 and other members of the PIKK family of proteins contain two conserved domains at the C-terminus, the PI3K and FATC domains. In other members of the PIKK family, the PI3K domain is a kinase and the FATC domain regulates the kinase activity of the adjacent PI3K domain. The PI3K and FATC domains of Tra1/TRRAP do not exhibit kinase activity, yet they are still important for Tra1 function. The FATC domains of mammalian ATM and DNA-PKcs interact with Tip60 in response to DNA damage (reviewed in Lempiainen and Halazonetis. 2009). Furthermore, the recruitment of yeast Mec1 to DNA damage sites is dependent on the interaction of the FATC domain of Mec1 with Rfa1 (Nakada et al. 2005). Therefore, it is possible that the role of the C-terminal domains of Tra1 is to interact with other proteins, thus regulating one or more activities of Tra1 or the activity of the interacting protein. In order to identify peptide/protein sequences with which Tra1 may interact, three experiments were performed: gel filtration chromatography, over-expressing the C-terminus of Tra1, and phage display. ***My aim was to identify whether the PI3K and FATC domains of Tra1 interact with other proteins.***

Gel filtration chromatography was performed to determine whether the FATC and PI3K domains of Tra1 stably associate with other proteins *in vivo*. Gel filtration separates proteins on the basis of size, while maintaining the integrity of the protein complexes. Therefore, if $Tra1_{PI3K/FATC}$ interacts with other proteins *in vivo*, $Tra1_{PI3K/FATC}$ expressed in yeast may fractionate differently than $Tra1_{PI3K/FATC}$ expressed in *E. coli*. Gel filtration

was a rapid method to explore whether further purification and mass spectrometry experiments should be performed using $Tra1_{PI3K/FATC}$.

Over expression of the PI3K and FATC domains of Tra1 was performed to determine whether the C-terminus of Tra1 interacts with and sequesters the transcriptional machinery. Over-expressing the strong transcriptional activator protein VP16 inhibits transcription in yeast because it sequesters the transcriptional machinery from endogenous cellular activators (Berger et al. 1992). If the C-terminus of Tra1 was toxic to the cell when over-expressed independently, it could be used as a tool to genetically isolate interacting proteins; that is, suppressor mutations could be identified that would provide clues into the identity of the interacting molecules.

Phage display was performed to identify potential peptide sequences that could interact with the PI3K and FATC domains of Tra1 ($Tra1_{PI3K/FATC}$). One advantage to using phage display is that a large number of potential peptide interactors can be simultaneously screened. Peptides are displayed on the phage surface, and are easily identified because their corresponding DNA sequences are encoded by the phage. These peptide interactors could provide unique insights on the molecular interactions of $Tra1_{PI3K/FATC}$. Furthermore, peptide inhibitors that destabilize Tra1 *in vivo* (similar to $tra1_{LA}$) could also arise, which would be useful in elucidating Tra1's function.

2. Results

2.1 Gel Filtration Chromatography

One role for the PI3K and FATC domains of Tra1 could be to interact with other proteins, thus regulating a function of Tra1 or a function of the interacting protein. As a first step to analyze whether the C-terminal domains of Tra1, the PI3K and FATC domains, interact with other proteins, I expressed the C-terminal 458 amino acid residues of Tra1 as a 9-Myc (Myc_9) and 6-histidine (6HIS) fusion in yeast (Myc_9-6HIS-$Tra1_{PI3K/FATC}$). If this protein interacts with other proteins it would fractionate with an apparent size greater than its actual size when analyzed by gel filtration chromatography. Sephadex® G-75 gel filtration beads, which exclude molecules larger than 70 kDa, were used in a spin column for this experiment.

To determine the migration profile of the C-terminus of Tra1 (residues 3290-3744) when not in complex with other proteins, I expressed the C-terminal 454 amino acid residues of Tra1 as a 6-histidine (6HIS) and general binding (GB) domain recombinant protein (6HIS-GB-$Tra1_{PI3K/FATC}$) in *E. coli* (strain CE4048). This molecule does not contain the Myc_9 tag but does contain the GB tag with a similar size (14 kDa). 6HIS-GB-$Tra1_{PI3K/FATC}$ expressed in *E. coli* primarily eluted from the G-75 centrifugation column in the fraction following the first wash (Figure 18 lower panel). This agrees with the fact that recombinant 6HIS-$Tra1_{PI3K/FATC}$ is ~65 kDa (smaller than 70 kDa).

Myc_9-6HIS-$Tra1_{PI3K/FATC}$ was analyzed on the same column. As shown in Figure 18, when residues 3286-3744 of Tra1 were expressed in yeast (on a multi-copy plasmid under control of the *MET3* promoter, strain CE5766) the protein was excluded from the G-75 centrifugation column, eluting both in the flow-through and first wash fractions.

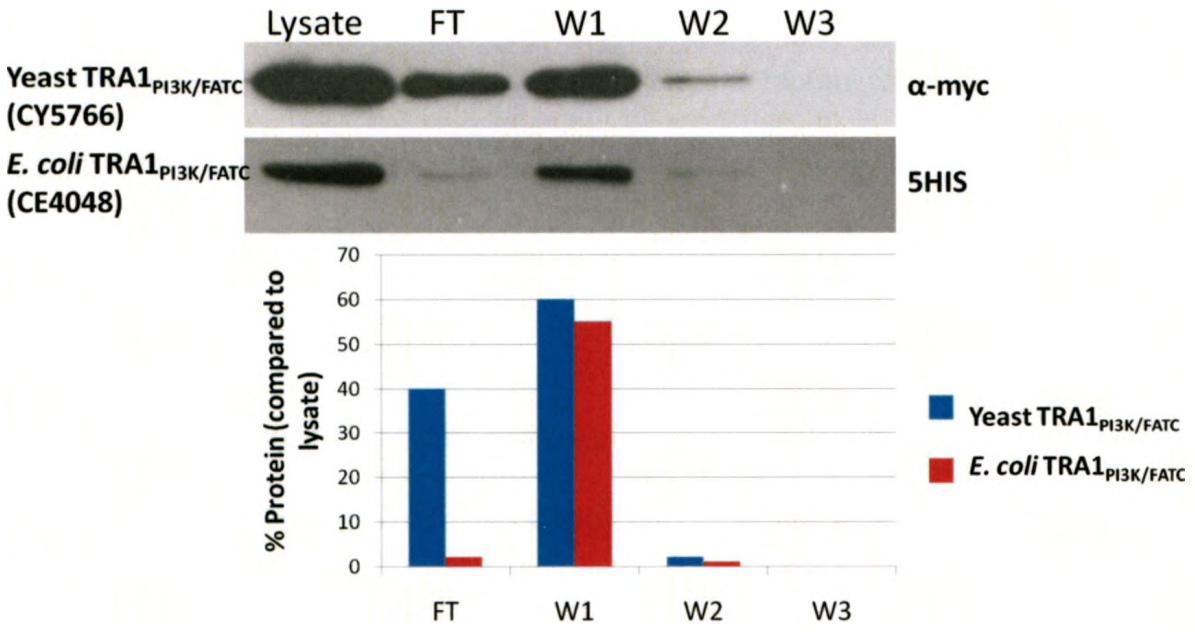

Figure 18: Size exclusion chromatography of Tra1$_{PI3K/FATC}$ expressed in yeast and *E. coli*. The C-terminus of Tra1 (Tra1$_{PI3K/FATC}$) was expressed both in *E. coli* CE4048 (6HIS-GB-Tra1$_{PI3K/FATC}$, residues 3290-3744) and in yeast CY5766 (Myc$_9$-6HIS-Tra1$_{PI3K/FATC}$, residues 3286-3744). Both constructs were ~65 kDa, expressing the C-terminus of Tra1. Cell lysates from each were analyzed on G-75 Sephadex spin columns. 0.5 mg of protein was loaded onto each column in a total volume of 100 µL of PO_4 Buffer 150. Columns were washed three-times (W1-W3) with 300 µL PO_4 Buffer 150. Twenty microlitres of each fraction was loaded onto two 10 % SDS-PAGE gels, transferred to PVDF membranes, and probed with the indicated antibodies. Band densities were quantified using the SpotDenso program of AlphaImager 3400. The histogram shows the band intensity of protein found in FT, W1, W2, and W3 for each sample as a percentage of that found in the lysate.

This suggests that Myc_9-6HIS-$Tra1_{PI3K/FATC}$ is likely interacting with one or more proteins in yeast. An obvious caveat of this experiment is that the two forms of the protein, though of very similar size (~14 kDa), are not identical. I can thus not exclude the possibility that the differences in migration are due to the different tags.

2.2 Over-expressing Myc_9-6HIS-$Tra1_{PI3K/FATC}$

Overexpression of transcriptional activators in yeast can result in toxicity, likely due to their sequestration of essential transcription factors. If toxicity is observed, it provides an excellent screening tool to genetically identify the interacting partners. The C-terminus of Tra1 (PI3K and FATC domains) was over-expressed to determine whether it was toxic to the cell. The yeast strain CY4455 contains a plasmid expressing residues 3286-3744 containing the PI3K and FATC domains of Tra1 (Myc_9-$Tra1_{PI3K/FATC}$) under control of the *MET3* promoter. Yeast strain CY2286 was also similarly constructed as a control, and contains a plasmid expressing Ngg1 (Myc_9-Ngg1). The *MET3* promoter is repressed in a dose-dependent manner by methionine (Mao et al. 2002). I thus analyzed whether over-expression of Myc_9-$Tra1_{PI3K/FATC}$ affected growth of cells by growing cells in the absence of methionine. Serial dilutions of cells expressing either Myc_9-$Tra1_{PI3K/FATC}$ or the Myc_9-Ngg1 control were plated onto minimal plates containing 0 μM, 5 μM, 10 μM, or 20 μM of methionine. As shown in Figure 19A, the cells grew comparably at every concentration of methionine, indicating that Myc_9-$Tra1_{PI3K/FATC}$, at levels that could be obtained from the *MET3* promoter, is not toxic to cells.

To verify that the Myc_9-$Tra1_{PI3K/FATC}$ was expressed, yeast strains CY4486 (Myc_9-Ngg1) and CY4455 (Myc_9-$Tra1_{PI3K/FATC}$) were grown in minimal media lacking

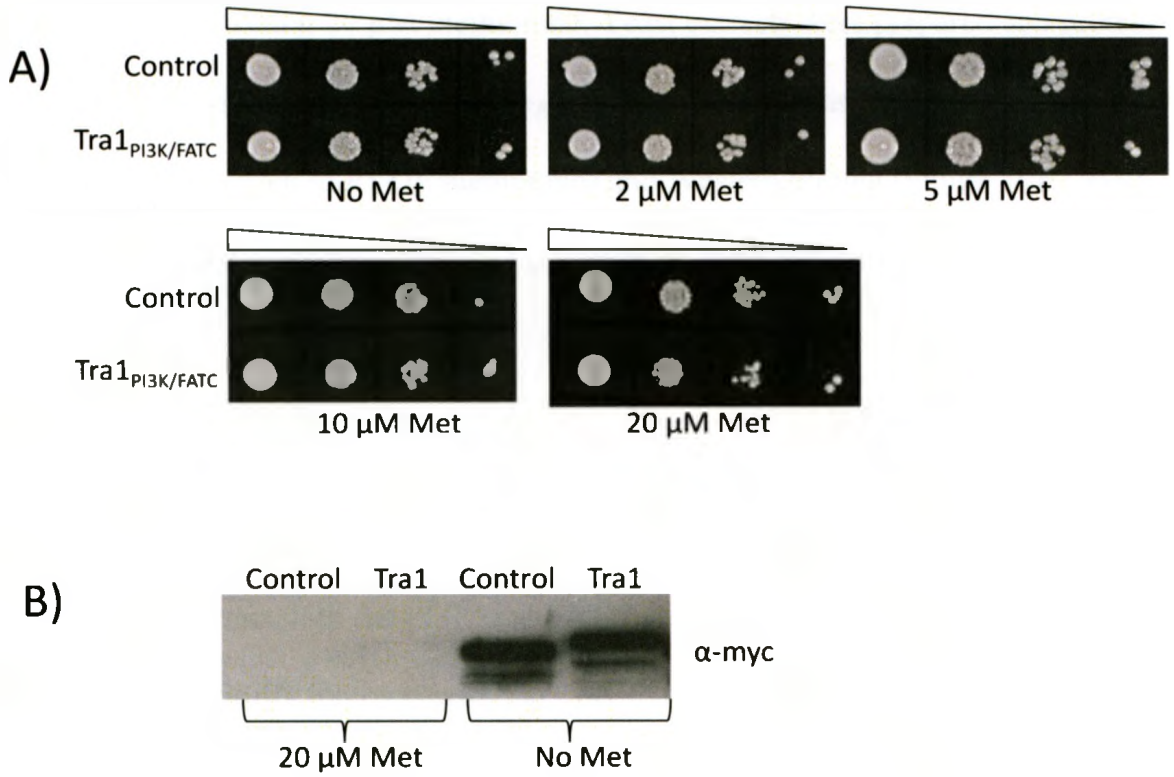

Figure 19: Over expression of the C-terminus of Tra1. A) Growth of strains containing plasmids expressing Myc_9-Ngg1 (Control) or the C-terminus of Tra1 from residue 3286 to 3744 (Myc_9-$Tra1_{PI3K/FATC}$) under the *MET3* promoter were compared on minimal plates containing the indicated concentrations of methionine. The level of expression decreases with increasing methionine (Mao, 2002). B) Myc_9-Ngg1 (Control) and Myc_9-$Tra1_{PI3K/FATC}$ (Tra1) were separated by SDS-PAGE on a 10 % gel, transferred to a PVDF membrane, western blotted, and probed with a primary α-myc antibody.

methionine or containing 20 μM methionine. Yeast extracts were prepared and protein extracts of cells with the Myc_9-$Tra1_{PI3K/FATC}$ or Myc_9-Ngg1 plasmids were separated by SDS-PAGE, transferred to PVDF membranes and Western blotted (Figure 19B). Both the Myc_9-Ngg1 control and Myc_9-$Tra1_{PI3K/FATC}$ proteins were expressed in the absence of methionine.

2.3 Phage Display

Phage display was performed to determine whether peptide sequences could be identified that interact with the PI3K and FATC domains of Tra1 ($Tra1_{PI3K/FATC}$). If a peptide that specifically interacts with the Tra1 C-terminus was discovered, it may provide clues as to which proteins interact with $Tra1_{PI3K/FATC}$. In addition these peptides could serve as inhibitors of Tra1 function.

The C-terminus of Tra1 (residues 3286-3744) was expressed as a recombinant protein (6HIS-GB-$Tra1_{PI3K/FATC}$) in *E. coli* (strain CE4048). A control recombinant protein (6HIS-GB) was also expressed in *E. coli* (strain CE4336). The PhD™-12 dodecapeptide M13 phage library (NEB), which covers 10^9 independent linear clones, was used in this experiment. The phage library was first incubated with 6HIS-GB control, to remove any phage that may bind to the IgG Sepharose or affinity tag. The phage were then bound to and eluted from the 6HIS-GB-$Tra1_{PI3K/FATC}$ column. Following amplification of phage, 5 rounds of panning were performed. Approximately ten-fold phage enrichment was observed after five rounds of panning, a result suggesting that specific phage had not been enriched. Normally, phage would be enriched by at least 1000 fold following only 3 rounds of panning (Duncan et al. 2011). In the final round, 5 plaques were sequenced and 4 of them contained the same sequence. When this peptide

was tested for specificity of 6HIS-GB-Tra1$_{PI3K/FATC}$, it bound equally to extracts from both the control 6HIS-GB and 6HIS-GB-Tra1$_{PI3K/FATC}$ (results not shown), strongly suggesting that it was not specific to 6HIS-GB-Tra1$_{PI3K/FATC}$.

To rule out trivial explanations for not isolating a specific interacting peptide, I examined whether phage was exposed to the target protein. The proteins used in the experiments (6HIS-GB and 6HIS-GB-Tra1$_{PI3K/FATC}$) were eluted from the IgG beads after each round of the experiment to ensure that protein was indeed present throughout the experiment (Figure 20). Protein bands corresponding to 20 kDa for 6HIS-GB and 65 kDa for 6HIS-GB-Tra1$_{PI3K/FATC}$ were observed in the elutions (Figure 20). Therefore, phage was exposed to target protein. Since one peptide bound to both 6HIS-GB and 6HIS-GB-Tra1$_{PI3K/FATC}$, it likely recognizes glycoprotein B since this domain was present in both.

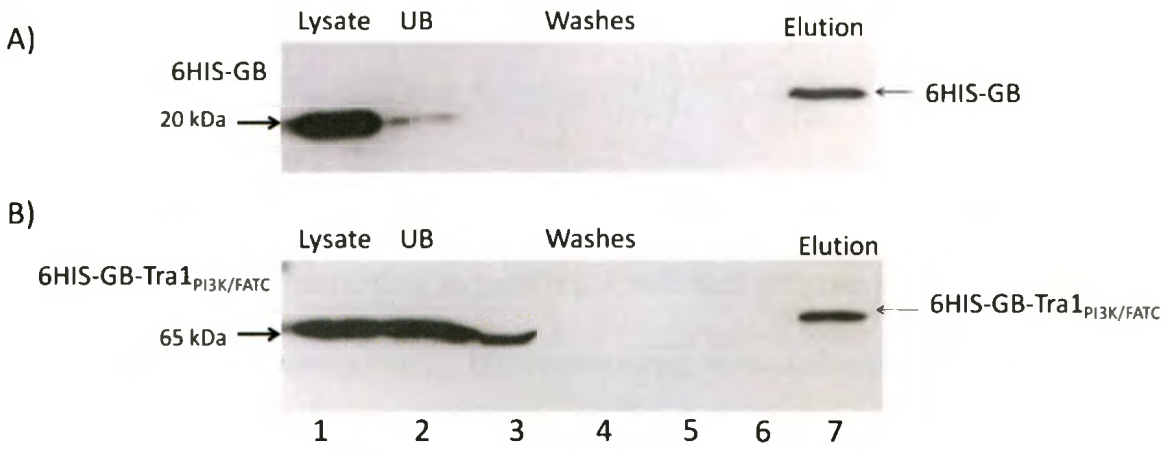

Figure 20: Proteins used for Phage display experiments. Crude extracts from *E. coli* containing plasmids expressing recombinant proteins of A) 6HIS-GB and B) 6HIS-GB-Tra1$_{PI3K/FATC}$, were applied to IgG sepharose beads. Proteins were eluted from the IgG beads with 1X Laemmli buffer (63 mM Tris HCl, 10 % Glycerol, 2 % SDS, 0.0025 % Bromophenol Blue, pH 6.8) and heating at 65 °C. Proteins were separated by SDS-PAGE on 15 % and 10 % gels respectively. Proteins were transferred to PVDF membranes, and western blotted probed with a 5HIS antibody to ensure that recombinant proteins 6HIS-GB and 6HIS-GB-Tra1$_{PI3K/FATC}$, were present throughout the phage display experiment. Lysates (*lane 1*), unbound protein (*lane 2*), washes (*lanes 3-6*) and eluted proteins (*lane 7* – 6HIS-GB and 6HIS-GB-Tra1$_{PI3K/FATC}$) are shown.

Chapter 5: Discussion

1. Prelude

This discussion will begin with an overview of characterizing suppressor mutations using next-generation sequencing. I will then propose several models for how *nam7Δ* may be suppressing $tra1_{LA}$. Differences in growth and suppression caused by variations in strain backgrounds will be discussed. I will discuss a possible model of how $tti2_{FS}$ may be suppressing $tra1_{FA}$, and how the FATC domain of Tra1 may be involved in protein interactions. Possible future directions that are currently being pursued will be encompassed throughout this discussion.

2. Identifying suppressor mutations using whole genome sequencing

Identifying genetic interactions is a powerful approach to provide insight into protein function. Particularly powerful is the approach of suppressor genetics which can reveal gene function, protein interactions, or parallel pathways in which the gene product is involved. In the suppressor genetics approach, the cell compensates for the cellular changes resulting from one genetic change by making a second suppressor mutation. The approach is particularly powerful because it requires no assumptions regarding the gene's function, and since the cell does the work it is not influenced by experimental bias. There are many examples in the literature that document the utility of suppressor genetics. For example, the *SPT* genes were originally identified by Fred Winston's laboratory as suppressors of the transcriptional inhibitory effects caused by Ty transposable elements (Winston et al. 1984). This suppression provided clues that Spt proteins had a role in

transcription. The *ADA* genes were first identified in the Guarente laboratory as suppressors of the toxicity of overexpression of the strong transcriptional activator VP16 (Berger et al. 1992). This suppression provided clues into the role of the Ada proteins in regulating transcription.

The initial identification of suppressor mutations through screens is relatively easy because the cells do the work. I performed a suppressor analysis using the $tra1_{LA}$ FATC domain mutant as the primary mutation. The CY4018 ($tra1_{LA}$) strain grows slowly on YPD plates containing 4 % ethanol. Therefore, to screen for suppressor mutations of $tra1_{LA}$, I plated the CY4018 strain on YPD plates containing 4 % ethanol and selected for colonies that were growing quickly. Eleven extragenic suppressor mutations were identified using this method. Random spore analysis was performed by mating suppressors with CY5522 ($tra1_{LA}$, *MATα*) and testing whether spore colonies grew quickly or were sensitive on YPD plates containing 4 % ethanol or grown at 37 °C. A 2:2 (slow growth on ethanol : fast growth on ethanol) segregation of spores indicates that the suppressor mutation is linked to a single gene. For *es2*, 27 spore colonies were analyzed; 13 grew quickly on YPD plates containing 4 % ethanol and 14 grew slowly, indicating a 2:2 segregation. Cumulatively, in the *es38* and *es41* complementation group, 42 spore colonies were analyzed; 19 grew quickly on YPD plates containing 4 % ethanol and 23 grew slowly, indicating an approximate 2:2 segregation. Had the complementation group been linked to more than one gene, 25% (~10 colonies out of 42) would have been expected to grow quickly on YPD plates containing 4 % ethanol. The next step in characterizing suppressor mutations involves determining whether the genes are dominant or recessive, and the number of complementation groups. Since yeast stably

exists as both haploid and diploid, determining complementation groups is straight forward. Through classical genetic approaches, I characterized the 11 $tra1_{LA}$ suppressor mutations as recessive and belonging to three complementation groups.

Until recently, the difficulty with suppressor genetics has come in the identification of the suppressor mutations. Methods for identifying the mutations have involved the addition of gene pools to revert the suppressed phenotype, or revert a novel phenotype (if it exists) caused by the suppressor. There are numerous difficulties with these traditional methods. For example, good gene pools are required, the plasmid may not fully revert the phenotype, and even after having isolated a reverting plasmid, the gene must be located. These traditional methods can be labour intensive and time consuming (Hobert. 2010). The low costs of whole-genome sequencing have overcome many of the difficulties in identifying the suppressor mutation.

Next generation sequencing is changing the landscape of genetics approaches because it allows for whole genome sequencing to be used to identify mutations (Hobert. 2010). Next-generation sequencing is a rapid and cost-effective approach to identifying mutations (Smith et al. 2008). Various laboratories have successfully used whole genome sequencing to identify mutations, and many more studies are underway (reviewed in Hobert. 2010).

The ABI (Applied Biosystems) SOLiD (Sequencing by Oligonucleotide Ligation Detection) 4.0 next-generation sequencing platform was used in our studies. ABI SOLiD is a next-generation sequencing approach that sequences using oligonucleotide ligation with dinucleotide probes. The DNA is fragmented and a library created. Single-stranded DNA fragments are attached to the surface of beads using adaptors. Each adaptor has a

unique barcode so that each sequence read contains a barcode. For paired-end reads, the library of DNA fragments is prepared and both ends of the fragments are sequenced with corresponding barcodes. Fluorescently labeled oligonucleotides are added, each specific for the identity of the two bases. These dinucleotides compete for incorporation into the DNA templates and the fluorescent signal is measured. Using this 2-base encoding approach, every single position or base in the DNA template is queried twice, thereby yielding a low base calling error (Chee-Seng et al. 2010). ABI SOLiD 4.0 has the lowest error rate (as of 2010) compared to the other next-generation sequencing platforms (Chee-Seng et al. 2010).

For my suppressor genetic analysis, I used a sequence alignment program, Bowtie, to map reads from each sample to the SGD reference genome. The SGD reference genome was used because it could easily be downloaded and converted to colourspace, therefore allowing mapping of the ABI SOLiD colourspace reads. The "v -3" parameter was used for Bowtie, which allows reads containing a maximum of three mismatches to be aligned to the SGD reference genome. This "v -3" parameter has been used in other studies that successfully identified point mutations (Smith et al. 2008). I also mapped paired-ends separately, rather than mapping them together as paired-ends. I did this so that reads without a paired-end were not eliminated from the analysis. In this manner, a higher percentage of total reads mapped to the genome, thereby maximizing the genome coverage.

An initial list of polymorphisms was obtained by eliminating those polymorphisms with a Phred score below 20. This means that polymorphisms present less than 99 % of the time were eliminated from the analysis. The $tra1_{FA}$ suppressor

mutations were originally identified in the consortium background (BY4741). The sequence of this strain background closely matches what is found in the SGD reference genome, so the initial list of filtered polymorphisms was short (there were ~100). Since the $tra1_{FA}$ pre-mutagenesis control was sequenced along with two $tra1_{FA}$ samples containing the same suppressor mutation, the background polymorphisms found in the $tra1_{FA}$ control were easily subtracted. Additionally, any polymorphisms that were not found in both of the suppressor mutation samples were eliminated. Once synonymous substitutions were eliminated, there was only one candidate suppressor for $tra1_{FA}$. This suppressor was a phenylalanine (F) to serine (S) mutation at amino acid residue 328 ($tti2_{FS}$). Therefore whole-genome sequencing combined with Bioinformatic analysis easily identified this suppressor.

The same approach and reasoning described above for the $tra1_{FA}$ mutation was used to identify the $tra1_{LA}$ suppressor mutations. However, the $tra1_{LA}$ suppressor mutations were identified in a different strain background (KY320) than $tra1_{FA}$. The sequence of this strain background does not so closely match what is found in the SGD reference strain. Therefore, the original list of polymorphisms was long (there were >3000), making the analysis more challenging. The background polymorphisms found in the $tra1_{LA}$ control sample were subtracted from each suppressor strain. Since $tra1_{LA}$ extragenic suppressor mutation *es2* was in a separate complementation group than *es38* and *es41*, those polymorphisms found in all three of these samples were also eliminated from the analysis (narrowing the polymorphism list down to <500). Synonymous substitutions were eliminated as before. Mutations in non-functional transposable elements, and highly repetitive intergenic regions were also eliminated. Once all these

background polymorphisms were eliminated, there was only one candidate suppressor for *es2*. This suppressor was a tryptophan (W) to opal mutation at amino acid 165 in *NAM7* ($nam7_{W\text{-}UM}$).

There was only one candidate suppressor for the *es38* and *es41* complementation group, a mutation in *srb6*. However, there was no coverage of this mutation in the $tra1_{LA}$ reference sample. After sequencing the *SRB6* gene, I found that it was in fact a background mutation also present in $tra1_{LA}$. There were no other candidate suppressors remaining for the *es38* and *es41* complementation group. There were no mutations present in RNA surveillance components Upf3 and Nmd2 in the *es38* and *es41* complementation group. Therefore, the identity of these suppressor mutations remains unknown. Given there was low coverage in some areas of the genome, it is possible that the suppressor mutation lies in an area of low coverage. A study sequencing the *Pichia pastoris* (methylotrophic yeast) genome using ABI SOLiD also found that some regions of the genome had more coverage than others (Smith et al. 2008). Since the "v -3" parameter in Bowtie was used, one limitation is that if the mutation were in a region with more than 3 polymorphisms nearby, the sequence would not have mapped to the SGD genome. However, 3 polymorphisms was the maximum number of mutations allowed using the Bowtie program so I could not increase this parameter. It may have been useful to build a reference genome using the $tra1_{LA}$ control, rather than using the SGD reference genome. However, it would have been extremely difficult to do this (Dr. Greg Gloor, personal communications) especially for a beginner Bioinformatician. Therefore, a future direction would be for someone with advanced knowledge in Bioinformatics to build a reference genome out of the $tra1_{LA}$ sequences. In this manner, the initial list of

polymorphisms would be low, heightening the chances of identifying the mutation of interest. Another reason why *es38* and *es41* may not have been identified is there was no method available using Bowtie to test whether the suppressor was an insertion or deletion. Dr. Gloor is looking into new programs, such as VarScan, that can identify insertions and deletions (Dr. Greg Gloor, personal communications).

ABI SOLiD 4.0 sequencing platform and my Bioinformatic analysis successfully identified $tti2_{FS}$ and $nam7_{W\text{-}UM}$ as the suppressor mutations of $tra1_{FA}$ and $tra1_{LA}$ respectively. Therefore, I have demonstrated that this method is a rapid and effective approach to identifying suppressor mutations.

3. It is unlikely that *nam7Δ* is increasing *TRA1* mRNA

One simple explanation for *nam7Δ* suppression of $tra1_{LA}$ is that *nam7Δ* is not targeting mRNAs for degradation and this is causing an increased stability of *TRA1* mRNA in the cell. mRNAs containing premature stop codons in *ACT1*, *PGK1*, *STE3*, and *MATα* were stabilized in *nam7Δ* strains (Leeds et al. 1991). Since the $tra1_{LA}$ and $tra1_{FA}$ mutants do not contain premature stop codons and the mutations are located very close to the C-terminus (ie: not upstream of DSE), this model seems unlikely. However, the NMD surveillance complex also targets regular mRNAs (Gardner. 2010) and mRNAs containing programmed frameshift mutations for NMD (Bidou et al. 2000). Hence, I cannot exclude the possibility that *TRA1* mRNA is a target of this pathway. My preliminary experiments suggest that there are no changes in *TRA1* mRNA levels between *TRA1*, $tra1_{LA}$, *nam7Δ* and a double $tra1_{LA}$ *nam7Δ*. This would suggest that suppression by *nam7Δ* is not by increasing *TRA1* mRNA levels. To obtain a more

definitive result, performing this experiment using real time PCR to observe a linear range and quantitatively confirm that there are no changes in *TRA1* RNA levels should be done.

4. Possible mechanisms for suppression of *tra1$_{LA}$* by *nam7Δ*

4.1 RNA's normally degraded by NMD may be compensating for lack of transcription due to *tra1$_{LA}$*

Changes in gene expression can occur in response to environmental stress. Phenotypes displayed by *tra1$_{LA}$* may occur because Tra1 is unable to regulate transcription of genes required in times of environmental stress (Hoke et al. 2010). Suppressor mutations may act by compensating for this change in gene expression. An example of such a model comes from the analysis of suppressor mutations of the deletion of the SAGA component *spt20*, which results in changes in histone acetylation and gene expression. The null version of the histone de-acetylase (HDAC) gene *hda1* is a suppressor of *spt20Δ* (Larschan and Winston. 2005). By deleting the HDAC *hda1*, the lack of HAT activity is compensated for and transcription is restored, at least at certain genes.

Gene expression is not only controlled by regulating transcription, but also through degradation of RNA transcripts. In fact, the RNA surveillance complex consisting of Nam7, Upf3, and Nmd2, regulates 5'decapping of both regular and nonsense mRNAs (Gardner. 2010, He and Jacobson. 2001). In this manner, Nam7 may regulate gene expression through RNA degradation (Gardner. 2010). Therefore, one model is that while *tra1$_{LA}$* is unable to regulate transcriptional activation of certain genes

required for the cellular stress response, *nam7Δ* (or *upf3Δ* or *nmd2Δ*) compensates at the RNA level by knocking out the NMD process and increasing the life of certain RNA transcripts (Figure 21).

The compensatory model would agree with *nam7Δ* partially restoring some of the growth phenotypes displayed by $tra1_{LA}$ because only processes that require genes regulated by NMD would be affected. To further test this model, expression profiling should be performed in a *nam7Δ* $tra1_{LA}$ background. These data could be compared to the expression profiling data that are already available for $tra1_{LA}$ (Hoke et al. 2010), to determine whether *nam7Δ* equilibrates RNAs with altered expression caused by $tra1_{LA}$.

4.2 Tra1 may be directly involved in NMD through regulating Nam7

Deletion of any of the three components of the NMD surveillance system (*nam7*, *nmd2*, and *upf3*) leads to nonsense-suppression phenotypes (Gonzalez et al. 2001). Since the $nam7_{W\text{-}UM}$ suppressor mutation was recessive, this suppressor is likely acting as a loss of function mutation. Indeed, I found that *nam7Δ*, *nmd2Δ*, and *upf3Δ* are suppressors of the slow growth on ethanol phenotype displayed by $tra1_{LA}$ suggesting that it is loss of the NMD machinery that suppresses $tra1_{LA}$. *nam7Δ* also suppressed Tra1 FATC domain mutant $tra1_{FA}$, but not SAGA or NuA4 components *ada2Δ*, *eaf3Δ* or *eaf7Δ*. Since *nam7Δ* suppresses Tra1 FATC domain mutants, this could indicate a role for Nam7 in regulating, or being regulated by, the FATC domain of Tra1.

In higher eukaryotes such as *C. elegans* and humans, SMG2, SMG3, and SMG4 are homologs of Nam7, Nmd2 and Upf3 respectively (Gonzalez et al. 2001). In these higher eukaryotes, SMG2 must be phosphorylated in order for NMD to take place

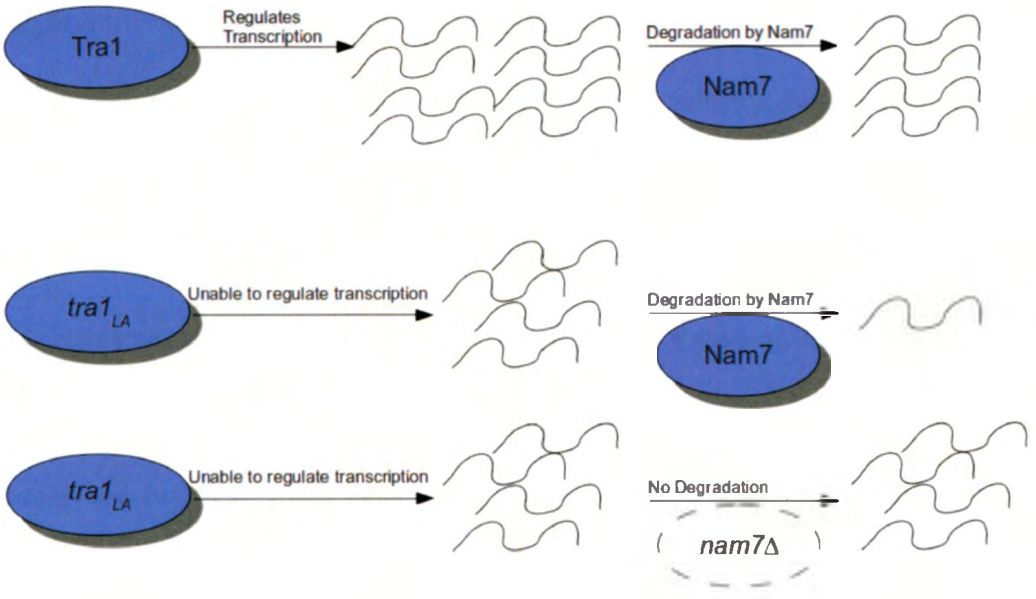

Figure 21: A model explaining compensatory suppression of *tra1*$_{LA}$ by *nam7Δ*. In times of stress, Tra1 regulates transcription of stress-response genes and Nam7 controls expression at the RNA level by degrading RNA transcripts. When *tra1*$_{LA}$ mutation occurs, certain stress-response genes are transcribed at a lower level, however, Nam7 degrades the RNA transcripts as normal. When *nam7Δ* occurs in the context of *tra1*$_{LA}$, these transcripts do not get degraded and the appropriate proteins can be translated in response to stress. This model would apply only to genes regulated both by Tra1 and Nam7.

(Wilusz et al. 2001). The PIKK protein SMG1, and two other proteins SMG3 and SMG4 are all required for the phosphorylation of SMG2 (Wilusz et al. 2001). Since SMG1 is a kinase and phosphorylates Nam7 *in vitro*, it is likely directly responsible for phosphorylating SMG2 (Grimson et al. 2004). Intriguingly, there is no yeast homolog of SMG1. Furthermore, Nam7 is not phosphorylated in yeast. Therefore, regulation of Nam7 activity must be different between yeast and higher eukaryotes (Wilusz et al. 2001). However, SMG1 is related to Tra1 in that it is part of the PIKK family. Therefore, it is possible Tra1 is working by a different mechanism than SMG1 to regulate Nam7. Although the C-terminus of Tra1 lacks kinase activity, it could still be regulating Nam7, either directly or through another protein.

Figure 22 shows a tentative model for regulation of Nam7 by Tra1. Perhaps the FATC domain is normally required for repressing activity of Nam7. When $tra1_{LA}$ occurs, the FATC domain of Tra1 is unable to repress Nam7. Unregulated Nam7 could then result in excessive degradation of mRNAs, consequently lowering protein levels and causing some of the stress-related phenotypes observed in $tra1_{LA}$ and $tra1_{FA}$. By deleting a component of the NMD machinery (such as *upf3Δ*, *nmd2Δ* or *nam7Δ*), NMD would be eliminated and some of the $tra1_{LA}$ phenotypes reversed. Further experiments are being performed to test whether Tra1 has a direct role in NMD. A *HIS4-LacZ* fusion of *PGK1* containing a premature stop codon upstream of the DSE, is currently being constructed in the Brandl laboratory. If Tra1 is directly involved in NMD, $tra1_{LA}$ should show increased read-through of the nonsense codon.

5. Differences in strain backgrounds cause growth differences

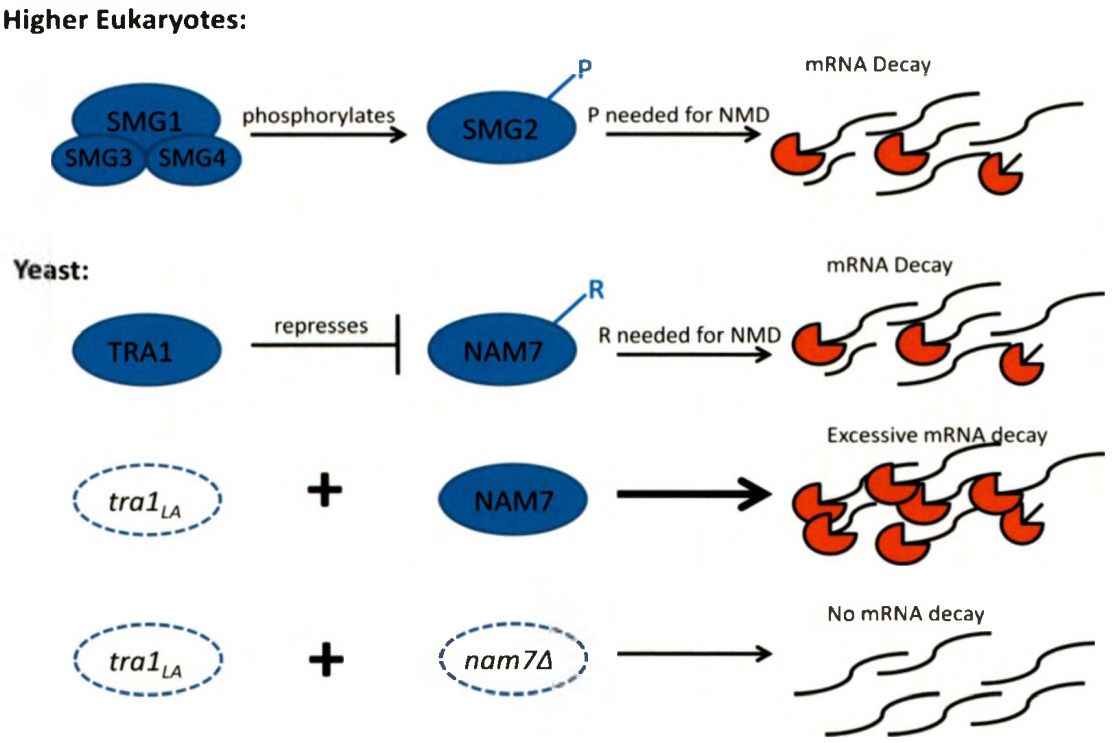

Figure 22: A model for Nam7 regulation by Tra1. In higher eukaryotes, SMG1 (along with SMG3 and SMG4) is required for phosphorylation of SMG2 (a Nam7 homolog). Phosphorylation (P) of SMG2 is essential for nonsense-mediated mRNA decay (NMD) to take place. There is no yeast homolog of SMG1 and Nam7 is not phosphorylated in yeast. Since Tra1 is structurally related to SMG1, I propose that Tra1 is repressing (R) Nam7 activity, thereby regulating NMD. In the mutant *tra1*$_{LA}$, perhaps this results in excessive NMD causing too much RNA degradation. When Nam7 is deleted (*nam7Δ*), the loss of NMD may restore balance to this pathway.

In the KY320 strain background (in which *tra1*$_{LA}$ is on a *CEN* plasmid), *PHO5* expression is reduced in the CY4018 (*tra1*$_{LA}$) strain, and partially restored in the strains containing suppressor mutations. However, in the BY4741 consortium background (in which *tra1*$_{LA}$ is integrated into the genome), *nam7Δ* does not restore *PHO5* expression (Dr. Chris Brandl, personal communication - results not shown). Furthermore, *nam7*$_{W\text{-}UM}$ in the KY320 strain partially restores the slow growth phenotype at 37 °C caused by *tra1*$_{LA}$, whereas *nam7Δ* in the consortium background does not. These differences in suppression may occur because *nam7*$_{W\text{-}UM}$ causes an increase in plasmid copy number thus increasing the level of *tra1*$_{LA}$ in the KY320 background. Therefore, experiments are currently being repeated using *nam7Δ* in the BY4741 background to determine which phenotypes displayed by *tra1*$_{LA}$ are suppressed by *nam7Δ*, and which phenotypes are due to strain background differences.

6. *nam7Δ* may be helping to stabilize *tra1*$_{LA}$ in the context of SAGA

In addition to *nam7Δ*, all 3 complementation groups of extragenic suppressor mutations partially restored growth of strains containing *tra1*$_{LA}$ on ethanol, tunicamycin, and tert-butylhydroperoxide, which are characteristic of SAGA mutants. Extragenic suppressor mutations did not restore growth phenotypes displayed by *tra1*$_{LA}$ on benomyl, rapamycin, or geneticin (results not shown), which are characteristic of NuA4 mutants. Therefore, perhaps the *tra1*$_{LA}$ extragenic suppressor mutations are suppressing Tra1's role within SAGA, more so than in NuA4. A high throughput screen identified the SAGA component Sgf29 as interacting with Nam7 (Gavin et al. 2006). Therefore, Sgf29 could be involved in linking Tra1 to a role in NMD in the context of SAGA.

7. The PI3K and FATC domains of Tra1 may be involved in protein-protein interactions

One proposed model for FATC domain function is that it may be required for protein interactions to take place. The hydrophobic residues found within the FATC domain of Tra1 resemble those residues found in proteins that interact with PDZ domains (Tonikian et al. 2008). Therefore, perhaps the FATC domain is required for protein interactions of Tra1. The FATC domain in other members of the PIKK family is required for kinase activity of the PI3K domain. Although Tra1 does not possess kinase activity, it may still play a regulatory role by promoting protein interactions at the C-terminus.

Gel filtration chromatography was performed to determine whether the PI3K and FATC domains of Tra1 interact with other proteins. The PI3K and FATC domains of Tra1 were expressed both in yeast (Myc_9-6HIS-$Tra1_{PI3K/FATC}$) and in *E. coli* (6HIS-GB-$Tra1_{PI3K/FATC}$). The GB domain is approximately the same size as the Myc_9 tag (14 kDa), and therefore both constructs were ~65 kDa. When expressed in yeast, Myc_9-6HIS-$Tra1_{PI3K/FATC}$ was excluded from the G-75 spin column, eluting primarily in the void and after the first wash. When expressed in *E. coli*, 6HIS-GB-$Tra1_{PI3K/FATC}$ eluted primarily after the first wash. This result suggests that the C-terminus of Tra1 may be interacting with a protein or protein complex. One limitation of this approach was that the tags on the two proteins, although the same size, were not identical. Hence, it is possible that differences in elution could be due to the different tags.

In light of the possibility that the C-terminal domain of Tra1 may be interacting with other proteins, experiments were performed that would allow for their identification. The PI3K and FATC domains of Tra1 were over-expressed to test whether this would be

toxic to the cell. If the C-terminus of Tra1 was toxic to the cell when over-expressed independently, it could be used as a tool to genetically isolate interacting proteins. Over-expressing the C-terminus of Tra1 was not toxic to the cell, and therefore this is not a possible manner to identify Tra1 FATC interacting proteins.

Phage display was also performed to identify potential peptide interactors. Phage display using Tra1 and related proteins has been successfully performed by other laboratories. One study identified novel interactions of HMG-1 using the PhD-12 phage display library (Dintilhac and Bernues. 2002). Another study used Tra1 as a target for identifying transcription factors with the potential as therapeutics for transcription related ailments (Majmudar et al. 2009). This study looked for peptide interactors of Tra1 (residues 3092-3524) using phage display and the same PhD-12 library. They identified one peptide that was able to bind to that region of Tra1, and to a Mediator component Med15 (Majmudar et al. 2009). I was exploring residues 3286-3744 of Tra1. Unfortunately, no peptide interactors specific to this region of Tra1 were identified. Reasons may include: lack of complexity in the phage library, simple peptides may not adopt a conformation able to bind the FATC domain, or the FATC domain may only interact with other molecules in the context of the intact Tra1 protein.

Suppression of $tra1_{FA}$ by $tti2_{FS}$ could arise by the direct interaction of these proteins as they are known to co-exist in the ASTRA complex (Shevchenko et al. 2008). Tti2 directly interacts with Tel2, a coordinator of PIKK activity (Takai et al. 2007) and another member of ASTRA. As an initial test to determine whether Tra1 and Tti2 were directly interacting, I constructed a fusion protein of Tti2 (residues 165-420) with a GST tag and expressed this construct in *E. coli*. Tti2 was immobilized onto a glutathione resin,

and CE4048 (C-terminus of Tra1, residues 3286-3744 in *E. coli*) was incubated with the resin to determine whether the recombinant proteins would interact. However, these regions of Tra1 and Tti2 did not interact (results not shown). The interaction between Tra1 and Tti2 in yeast is currently being investigated in the Brandl laboratory. So far, immunoprecipitations have not demonstrated a strong interaction between wild-type Tra1 and Tti2; however, $tra1_{FA}$ interacted weakly with Tti2. Therefore, perhaps $tra1_{FA}$ is not able to release Tti2, and the $tti2_{FS}$ mutation will reverse this effect. These models are preliminary and are currently being investigated further.

Chapter 6: Conclusions

Suppressor genetics is a powerful method for identifying new interactions or protein regulatory functions. My research identified $nam7_{W\text{-}UM}$ as a suppressor of $tra1_{LA}$. I demonstrated that deletions of *nam7*, *nmd2*, and *upf3* are suppressors of Tra1 FATC domain mutants $tra1_{LA}$ and $tra1_{FA}$. My bioinformatics analysis also identified the $tti2_{FS}$ suppressor of $tra1_{FA}$. Allele specificity tests showed that these suppressors do not suppress other SAGA or NuA4 mutants, suggesting that they specifically target Tra1 function.

Nam7 is part of the surveillance complex mediating NMD. In higher eukaryotes, the PIKK protein SMG1 is required for the phosphorylation of SMG2 (Nam7 homolog). Phosphorylation of SMG2 is required for NMD to occur. Since Tra1 is structurally related to SMG1, perhaps it is also regulating Nam7 in the NMD pathway. Alternatively, lack of NMD may be compensating for transcriptional defects caused by $tra1_{LA}$ by not degrading RNA transcripts. Further experiments are currently underway to test whether Tra1 has a direct role in NMD.

Tra1 is a member of three major protein complexes: SAGA, NuA4, and ASTRA. It is a member of the PIKK family of proteins which are important cellular regulators. Although the function of the FATC domain of Tra1 is unknown, it may be required for protein interactions to take place. My gel filtration results suggest that Tra1 is interacting with other molecules within the cell. The results of this thesis have linked Tra1 to a new cellular process – NMD, and further connected Tra1 to the ASTRA complex. Deciphering the FATC domain function of Tra1 may be a key in connecting Tra1 with new regulatory processes.

References:

Abraham RT (2004) PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways. DNA Repair (Amst) 3:883-7.

Ahmed M and Rahman N (2006) ATM and breast cancer susceptibility. Oncogene 25:5906-5911.

Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2010) Essential cell biology, 3rd edition. Garland Science, New York, NY, USA.

Allard S, Utley RT, Savard J, Clarke A, Grant P, Brandl CJ, Pillus L, Workman JL, Côté J (1999) NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J. 18:5108-5119.

Auger A, Galarneau L, Altaf M, Nourani A, Doyon Y, Utley RT, Cronier D, Allard S, Côté J (2008) Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants. Mol.Cell.Biol. 28:2257-2270.

Baker SP and Grant PA (2007) The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. Oncogene 26:5329-5340.

Beamish HJ, Jessberger R, Riballo E, Priestley A, Blunt T, Kysela B, Jeggo PA (2000) The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. Nucleic Acids Res. 28:1506-13.

Berger SL, Pina B, Silverman N, Marcus GA, Agapite J, Regier JL, Triezenberg SJ, Guarente L (1992) Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. Cell 70:251-65.

Bhaumik SR and Green MR (2001) SAGA is an essential in vivo target of the yeast acidic activator Gal4p. Genes Dev. 15:1935-1945.

Bhaumik SR, Raha T, Aiello DP, Green MR (2004) In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes Dev. 18:333-343.

Bidou L, Hatin I, Perez N, Allamand V, Panthier JJ, Rousset JP (2004) Premature stop codons involved in muscular dystrophies show a broad spectrum of readthrough efficiencies in response to gentamicin treatment. Gene Ther. 11:619-627.

Bidou L, Stahl G, Hatin I, Namy O, Rousset JP, Farabaugh PJ (2000) Nonsense-mediated decay mutants do not affect programmed -1 frameshifting. RNA 6:952-961.

Bird AW, Yu DY, Pray-Grant M, Qiu Q, Harmon KE, Megee PC, Grant PA, Smith MM, Christman MF (2002) Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. Nature 419:411-415.

Bosotti R, Isacchi A, Sonnhammer EL (2000) FAT: a novel domain in PIK-related kinases. Trends Biochem.Sci. 25:225-227.

Brandl CJ, Furlanetto AM, Martens JA, Hamilton KS (1993) Characterization of NGG1, a novel yeast gene required for glucose repression of GAL4p-regulated transcription. EMBO J. 12:5255-65.

Brown CE, Howe L, Sousa K, Alley SC, Carrozza MJ, Tan S, Workman JL (2001) Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. Science 292:2333-2337.

Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84:843-851.

Burgess RJ, Zhou H, Han J, Zhang Z (2010) A role for Gcn5 in replication-coupled nucleosome assembly. Mol.Cell 37:469-480.

Cairns BR (2009) The logic of chromatin architecture and remodelling at promoters. Nature 461:193-198.

Chee-Seng K, En Yun L, Yudi P, Kee-Seng C (2010) eLS; Next Generation Sequencing Technologies and Their Applications.

Chen W, and Struhl K (1988) Saturation mutagenesis of a yeast his3 "TATA element": genetic evidence for a specific TATA-binding protein Proc Natl Acad Sci 85: 2691-2695.

Chomczynski P and Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal.Biochem. 162:156-159.

Cui Y, Hagan KW, Zhang S, Peltz SW (1995) Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes Dev. 9:423-36.

Dames SA, Mulet JM, Rathgeb-Szabo K, Hall MN, Grzesiek S (2005) The solution structure of the FATC domain of the protein kinase target of rapamycin suggests a role for redox-dependent structural and cellular stability. J.Biol.Chem. 280:20558-20564.

Daniel JA and Grant PA (2007) Multi-tasking on chromatin with the SAGA coactivator complexes. Mutat.Res. 618:135-148.

Dintilhac A and Bernues J (2002) HMGB1 interacts with many apparently unrelated proteins by recognizing short amino acid sequences. J.Biol.Chem. 277:7021-7028.

Duncan KE, Dempsey BR, Killip LE, Adams J, Bailey ML, Lajoie GA, Litchfield DW, Brandl CJ, Shaw GS, Shilton BH (2011) Discovery and characterization of a nonphosphorylated cyclic peptide inhibitor of the peptidylprolyl isomerase, Pin1. J.Med.Chem. 54:3854-3865.

Fishburn J, Mohibullah N, Hahn S (2005) Function of a eukaryotic transcription activator during the transcription cycle. Mol.Cell 18:369-378.

Gardner LB (2010) Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis. Mol Cancer Res 8:295-308.

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

Gietz RD and Sugino A (1988) New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. Gene 74:527-534.

Gietz RD and Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350:87-96.

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

Gonzalez CI, Bhattacharya A, Wang W, Peltz SW (2001) Nonsense-mediated mRNA decay in Saccharomyces cerevisiae. Gene 274:15-25.

Grant PA, Duggan L, Cote J, Roberts SM, Brownell JE, Candau R, Ohba R, Owen-Hughes T, Allis CD, Winston F, Berger SL, Workman JL (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev. 11:1640-50.

Grant PA, Schieltz D, Pray-Grant M, Yates JR, Workman JL (1998) The ATM-related cofactor Tra1 is a component of the purified SAGA complex. Mol.Cell 2:863-867.

Green MR (2005) Eukaryotic transcription activation: right on target. Mol.Cell 18:399-402.

Grimson A, O'Connor S, Newman CL, Anderson P (2004) SMG1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. Mol.Cell.Biol. 24:7483-7490.

Hahn S, Buratowski S, Sharp PA, Guarente L (1989) Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the SPT15 suppressor of Ty element insertions. Cell 58:1173-1181.

Han M, Kim UJ, Kayne P, Grunstein M (1988) Depletion of histone H4 and nucleosomes activates the PHO5 gene in Saccharomyces cerevisiae. EMBO J. 7:2221-8.

He F and Jacobson A (2001) Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. Mol.Cell.Biol. 21:1515-30.

Helmlinger D, Marguerat S, Villen J, Swaney DL, Gygi SP, Bahler J, Winston F (2011) Tra1 has specific regulatory roles, rather than global functions, within the SAGA co-activator complex. EMBO J. 30:2843-2852.

Herceg Z, Hulla W, Gell D, Cuenin C, Lleonart M, Jackson S, Wang ZQ (2001) Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression. Nat.Genet. 29:206-211.

Hobert O (2010) The impact of whole genome sequencing on model system genetics: get ready for the ride. Genetics 184:317-319.

Hoke SM, Irina Mutiu A, Genereaux J, Kvas S, Buck M, Yu M, Gloor GB, Brandl CJ (2010) Mutational analysis of the C-terminal FATC domain of Saccharomyces cerevisiae Tra1. Curr.Genet. 56:447-465.

Horejsi Z, Takai H, Adelman CA, Collis SJ, Flynn H, Maslen S, Skehel JM, de Lange T, Boulton SJ (2010) CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability. Mol.Cell 39:839-850.

Huisinga KL and Pugh BF (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Mol.Cell 13:573-585.

Hurov KE, Cotta-Ramusino C, Elledge SJ (2010) A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. Genes Dev. 24:1939-1950.

Imoberdorf RM, Topalidou I, Strubin M (2006) A role for gcn5-mediated global histone acetylation in transcriptional regulation. Mol.Cell.Biol. 26:1610-1616.

Jacobson S and Pillus L (2009) The SAGA subunit Ada2 functions in transcriptional silencing. Mol.Cell.Biol. 29:6033-45.

Johansson MJ and Jacobson A (2010) Nonsense-mediated mRNA decay maintains translational fidelity by limiting magnesium uptake. Genes Dev. 24:1491-1495.

Jonsson ZO, Jha S, Wohlschlegel JA, Dutta A (2004) Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. Mol.Cell 16:465-477.

Kanoh J and Yanagida M (2007) Tel2: a common partner of PIK-related kinases and a link between DNA checkpoint and nutritional response? Genes Cells 12:1301-1304.

Keith CT and Schreiber SL (1995) PIK-related kinases: DNA repair, recombination, and cell cycle checkpoints. Science 270:50-1.

Keogh MC, Mennella TA, Sawa C, Berthelet S, Krogan NJ, Wolek A, Podolny V, Carpenter LR, Greenblatt JF, Baetz K, Buratowski S (2006) The Saccharomyces cerevisiae histone H2A variant Htz1 is acetylated by NuA4. Genes Dev. 20:660-665.

Koutelou E, Hirsch CL, Dent SY (2010) Multiple faces of the SAGA complex. Curr.Opin.Cell Biol. 22:374-382.

Kouzarides T (2002) Histone methylation in transcriptional control. Curr.Opin.Genet.Dev. 12:198-209.

Kurabe N, Katagiri K, Komiya Y, Ito R, Sugiyama A, Kawasaki Y, Tashiro F (2007) Deregulated expression of a novel component of TFTC/STAGA histone acetyltransferase complexes, rat SGF29, in hepatocellular carcinoma: possible implication for the oncogenic potential of c-Myc. Oncogene 26:5626-5634.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10(3):R25

Larschan E and Winston F (2005) The Saccharomyces cerevisiae Srb8-Srb11 complex functions with the SAGA complex during Gal4-activated transcription. Mol.Cell.Biol. 25:114-123.

Lee TI, Causton HC, Holstege FC, Shen WC, Hannett N, Jennings EG, Winston F, Green MR, Young RA (2000) Redundant roles for the TFIID and SAGA complexes in global transcription. Nature 405:701-704.

Leeds P, Peltz SW, Jacobson A, Culbertson MR (1991) The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. Genes Dev. 5:2303-2314.

Lempiainen H and Halazonetis TD (2009) Emerging common themes in regulation of PIKKs and PI3Ks. EMBO J. 28:3067-73.

Lu PY, Lévesque N, Kobor MS (2009) NuA4 and SWR1-C: two chromatin-modifying complexes with overlapping functions and components. Biochem.Cell Biol. 87:799-815.

Majmudar CY, Labut AE, Mapp AK (2009) Tra1 as a screening target for transcriptional activation domain discovery. Bioorg.Med.Chem.Lett. 19:3733-3735.

Mao X, Hu Y, Liang C, Lu C (2002) MET3 promoter: a tightly regulated promoter and its application in construction of conditional lethal strain. Curr.Microbiol. 45:37-40.

Martinez E (2002) Multi-protein complexes in eukaryotic gene transcription. Plant Mol.Biol. 50:925-947.

McMahon SB, Van Buskirk ,H.A., Dugan KA, Copeland TD, Cole MD (1998) The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. Cell 94:363-374.

Mitchell L, Lambert JP, Gerdes M, Al-Madhoun AS, Skerjanc IS, Figeys D, Baetz K (2008) Functional dissection of the NuA4 histone acetyltransferase reveals its role as a genetic hub and that Eaf1 is essential for complex integrity. Mol.Cell.Biol. 28:2244-2256.

Mitchell P and Tollervey D (2003) An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. Mol.Cell 11:1405-1413.

Morita T, Yamashita A, Kashima I, Ogata K, Ishiura S, Ohno S (2007) Distant N- and C-terminal domains are required for intrinsic kinase activity of SMG1, a critical component of nonsense-mediated mRNA decay. J.Biol.Chem. 282:7799-7808.

Murr R, Vaissiere T, Sawan C, Shukla V, Herceg Z (2007) Orchestration of chromatin-based processes: mind the TRRAP. Oncogene 26:5358-5372.

Mutiu AI and Brandl CJ (2005) RNA isolation from yeast using silica matrices. J Biomol Tech 16:316-317.

Mutiu AI, Hoke SM, Genereaux J, Hannam C, MacKenzie K, Jobin-Robitaille O, Guzzo J, C\ôté J, Andrews B, Haniford DB, Brandl CJ (2007) Structure/function analysis of the phosphatidylinositol-3-kinase domain of yeast tra1. Genetics 177:151-166.

Naar AM, Lemon BD, Tjian R (2001) Transcriptional coactivator complexes. Annu.Rev.Biochem. 70:475-501.

Nakada D, Hirano Y, Tanaka Y, Sugimoto K (2005) Role of the C terminus of Mec1 checkpoint kinase in its localization to sites of DNA damage. Mol.Biol.Cell 16:5227-5235.

Narlikar GJ, Fan HY, Kingston RE (2002) Cooperation between complexes that regulate chromatin structure and transcription. Cell 108:475-487.

Park J, Kunjibettu S, McMahon SB, Cole MD (2001) The ATM-related domain of TRRAP is required for histone acetyltransferase recruitment and Myc-dependent oncogenesis. Genes Dev. 15:1619-24.

Priestley A, Beamish HJ, Gell D, Amatucci AG, Muhlmann-Diaz M, Singleton BK, Smith GC, Blunt T, Schalkwyk LC, Bedford JS, Jackson SP, Jeggo PA, Taccioli GE (1998) Molecular and biochemical characterisation of DNA-dependent protein kinase-defective rodent mutant irs-20. Nucleic Acids Res. 26:1965-1973.

Reeves WM and Hahn S (2005) Targets of the Gal4 transcription activator in functional transcription complexes. Mol.Cell.Biol. 25:9092-9102.

Rodriguez-Navarro S (2009) Insights into SAGA function during gene expression. EMBO Rep. 10:843-50.

Runge KW and Zakian VA (1996) TEL2, an essential gene required for telomere length regulation and telomere position effect in Saccharomyces cerevisiae. Mol.Cell.Biol. 16:3094-3105.

Saleh A, Schieltz D, Ting N, McMahon SB, Litchfield DW, Yates JR, Lees-Miller S, Cole MD, Brandl CJ (1998) Tra1p is a component of the yeast Ada.Spt transcriptional regulatory complexes. J.Biol.Chem. 273:26559-26565.

Shevchenko A, Roguev A, Schaft D, Buchanan L, Habermann B, Sakalar C, Thomas H, Krogan NJ, Shevchenko A, Stewart AF (2008) Chromatin Central: towards the comparative proteome by accurate mapping of the yeast proteomic environment. Genome Biol. 9(11):R167

Sibanda BL, Chirgadze DY, Blundell TL (2010) Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. Nature 463:118-121.

Smith DR, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, Shen L, Donahue WF, Tusneem N, Stromberg MP, Stewart DA, Zhang L, Ranade SS, Warner JB, Lee CC, Coleman BE, Zhang Z, McLaughlin SF, Malek JA, Sorenson JM, Blanchard AP, Chapman J, Hillman D, Chen F, Rokhsar DS, McKernan KJ, Jeffries TW, Marth GT, Richardson PM (2008) Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Res. 18:1638-1642.

Spagnolo L, Rivera-Calzada A, Pearl LH, Llorca O (2006) Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. Mol.Cell 22:511-519.

Strahl BD and Allis CD (2000) The language of covalent histone modifications. Nature 403:41-5.

Sun Y, Jiang X, Chen S, Fernandes N, Price BD (2005) A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. PNAS 102:13182-7.

Takahashi T, Hara K, Inoue H, Kawa Y, Tokunaga C, Hidayat S, Yoshino K, Kuroda Y, Yonezawa K (2000) Carboxyl-terminal region conserved among phosphoinositide-kinase-related kinases is indispensable for mTOR function in vivo and in vitro. Genes Cells 5:765-75.

Takai H, Wang RC, Takai KK, Yang H, de Lange T (2007) Tel2 regulates the stability of PI3K-related protein kinases. Cell 131:1248-1259.

Takai H, Xie Y, de Lange T, Pavletich NP (2010) Tel2 structure and function in the Hsp90-dependent maturation of mTOR and ATR complexes. Genes Dev. 24:2019-2030.

Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science 294: 2364-2368.

Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh J, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin X, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS (2008) A specificity map for the PDZ domain family. PLoS Biol 6:e239.

Villard J (2004) Transcription regulation and human diseases. Swiss Med.Wkly. 134:571-9.

Weake VM and Workman JL (2008) Histone ubiquitination: triggering gene activity. Mol.Cell 29:653-63.

Wilusz CJ, Wang W, Peltz SW (2001) Curbing the nonsense: the activation and regulation of mRNA surveillance. Genes Dev. 15:2781-2785.

Winston F, Chaleff DT, Valent B, Fink GR (1984) Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics 107:179-197.

Winston F, Dollard C, Malone EA, Clare J, Kapakos JG, Farabaugh P, Minehart PL (1987) Three genes are required for trans-activation of Ty transcription in yeast. Genetics 115:649-656.

Winzeler EA and Davis RW (1997) Functional analysis of the yeast genome. Curr.Opin.Genet.Dev. 7:771-776.

Wu PY, Ruhlmann C, Winston F, Schultz P (2004) Molecular architecture of the S. cerevisiae SAGA complex. Mol.Cell 15:199-208.

Wu PY and Winston F (2002) Analysis of Spt7 function in the Saccharomyces cerevisiae SAGA coactivator complex. Mol.Cell.Biol. 22:5367-5379.

Zhang S, Ruiz-Echevarria M, Quan Y, Peltz SW (1995) Identification and characterization of a sequence motif involved in nonsense-mediated mRNA decay. Mol.Cell.Biol. 15:2231-2244.

Appendix:

Table S1: Yeast and *E.coli* strains

Strain Number	Description	Plasmids	Reference
KY320	*MATa ura3-52 ade2-101 trp1-Δ1 lys2-801 his3-Δ200 leu2::PET56*		Chen, 1988
CY2706	Isogenic to KY320 except *tra1::Tn10LUK*	*YCplac22-myc$_9$-TRA1*	Hoke, 2010
CY4018	Isogenic to CY2706	*YCplac22u'-myc$_9$-tra1-L3733A*	Hoke, 2010
CY3003	Isogenic to CY2706	*YCplac22-myc$_9$-TRA1*	Hoke, 2010
CY4413	Isogenic to KY320		This study
CY5522	Isogenic to CY3003 except *MATα*	*YCplac22-myc$_9$-TRA1*	This study
CY5579	*es2* (*nam7$_{W\text{-}UM}$*), otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5580	*es12*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5581	*es35*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5582	*es36*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5583	*es37*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5584	*es38*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5585	*es39*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5586	*es40*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5587	*es41*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5588	*es42*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5750	*es43*, otherwise isogenic to CY4018	*YCplac22u-myc$_9$-tra1-L3733A*	This study
CY5666	Isogenic to CY5579 except *MATα*	*YCplac22-myc$_9$-tra1-L3733A*	This study
CY5758	Isogenic to CY5584 except *MATα*	*YCplac22-myc$_9$-tra1-L3733A*	This study
CY5603	Isogenic to CY5587 except *MATα*	*YCplac22-myc$_9$-tra1-L3733A*	This study
CY5688	Isogenic to CY5579	*YCplac22-myc$_9$-TRA1*	This study
CY5690	Isogenic to CY5584	*YCplac22-myc$_9$-TRA1*	This study
CY5691	Isogenic to CY5587	*YCplac22-myc$_9$-TRA1*	This study
CE4336	ER2566 (NEB strain)	6HIS-GB-pET21a	This study
CE4048	ER2566 (NEB strain)	6HIS-GB-Tra1$_{PI3K/FATC}$-pET21a	This study
ER2738			NEB strain
BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*		Winzeler, 1997
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*		Winzeler, 1997
BY 4743	*MATa/α his3Δ1/his3Δ1 leu2Δ0 /leu2Δ0 lys2Δ0/LYS2 MET15/met15Δ0 ura3Δ0 /ura3Δ0*		Winzeler, 1997
CY4486	Isogenic to BY4742	pS14 (Myc$_9$-Ngg1)	This study
CY4455	Isogenic to BY4742	pS17(Myc$_9$-Tra1$_{PI3K/FATC}$)	This study
CY5766	Isogenic to BY4742	pS26 (Myc$_9$-6HIS-Tra1$_{PI3K/FATC}$)	This study
BY46214	*MATa/α his3Δ1/his3Δ1 leu2Δ0 lys2Δ0/LYS2 MET15/met15Δ0 ura3Δ0/ura3Δ0 + can1Δ::LEU2+-MFA1pr-HIS3/CAN1 + nam7Δ*::kan^R		Tong et al., 2001
CY5932	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nam7Δ*::kan^R		This study
CY5933	Isogenic to CY5932 except *MATα*		This study
CY5938	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nam7Δ*::nat^R		This study

CY5939	Isogenic to CY5938 except *MATα*		This study
BY41905	*MAT*a/α *his3Δ1/his3Δ1 leu2Δ0 lys2Δ0/LYS2 MET15/met15Δ0 ura3Δ0/ura3Δ0 + can1Δ::LEU2+-MFA1pr-HIS3/CAN1+ nmd2Δ*::kan^R		Tong et al., 1997
CY5934	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nmd2Δ*::kan^R		This study
CY5935	Isogenic to CY5934 except *MATα*		This study
BY44702	*MAT*a/α *his3Δ1/his3Δ1 leu2Δ0 lys2Δ0/LYS2 MET15/met15Δ0 ura3Δ0/ura3Δ0 + can1Δ::LEU2+-MFA1pr-HIS3/CAN1+ upf3Δ*::kan^R		Tong et al, 1997
CY5936	*MATa HIS3 leu2Δ0 met15Δ0 ura3Δ0 upf3Δ*::kan^R		This study
CY5937	Isogenic to CY5936 except *MATα* and *his3Δ*		This study
CY5974	*MATα ura3Δ0 LEU2 met15Δ0 tra1-F3744A-HIS3 nam7Δ*::kan^R	*YHR100C* in *Ycplac33*	This study
CY5976	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 eaf7*::kan^R *nam7Δ*::nat^R		This study
CY5979	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ada2*::kan^R *nam7Δ*::nat^R		This study
CY5980	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 eaf3*::kan^R *nam7Δ*::nat^R		This study
CY5983	*MATα ura3Δ0 leu2Δ0 met15Δ0 tra1-L3733A-HIS3 upf3Δ*::kan^R	*YHR100C* in *Ycplac33*	This study
CY5972	*MATα ura3Δ0 leu2Δ0 met15Δ0 tra1-L3733A-HIS3 nam7Δ*::kan^R	*YHR100C* in *Ycplac33*	This study
CY5996	*MATα ura3Δ0 leu2Δ0 met15Δ0 tra1-L3733A-HIS3 nmd2Δ*::kan^R	*YHR100C* in *Ycplac33*	This study
CY4434	*MATa ura3Δ0 his3Δ0 leu2Δ0 TRA1-HIS3*	*YHR100C* in *Ycplac33*	This study
BY4282	Isogenic to BY4741 except *ada2*::kan^R		Winzeler, 1997
BY7143	Isogenic to BY4741 except *eaf3*::kan^R		Winzeler, 1997
BY2940	Isogenic to BY4741 except *eaf7*::kan^R		Winzeler, 1997
CY5967	*MATα ura3Δ0 his3Δ0 leu2Δ0 tra1-L3733A-HIS3*	*YHR100C* in *Ycplac33*	This study
CY4400	*MATa ura3Δ0 his3Δ0 leu2Δ0 tra1-F3744A-HIS3*	*YHR100C* in *Ycplac33*	This study
CY5684	Isogenic to CY4400 except *tti2*-F328S	*YHR100C* in *Ycplac33*	This study
CY5685	Isogenic to CY5684	*YHR100C* in *Ycplac33*	This study
CY5686	Isogenic to CY4400	*YHR100C* in *Ycplac33*	This study

[1] *YCplac22u*: *YCplac22* except *TRP1* marker has been switched for *URA3*

Table S2: Oligonucleotides used in this study

Name	Sequence (5' to 3')	Description
4292-1	ATGATCAATAACGCTACT CCTCACAAAATCC	Forward primer to verify $tra1_{LA}$
5632-1	CACGTCTGTTCTTGCTGC	Reverse primer to verify $tra1_{LA}$
5759-1	ATAAGAATGCGGCCGCTT TGTTGGCGCCTTATATAA G	Forward primer to amplify PI3K and FATC domains of *TRA1*
4479-1	ATACGAGCTCTTTGAGGC TTTCTCTACCTTC	Reverse primer to amplify 3' end of *TRA1*
1880	GGCCCATCACCATCATCA CCATGC	Oligonucleotide to make 6HIS tag
1881	GGCCGCATGGTGATGATG GTGATG	Oligonucleotide to make 6HIS tag
6076-1	ACTGTGGACGTAAGAACG TG	Forward primer to confirm $nam7_{W\text{-}UM}$
6076-2	TGGATCTCCATTTTGCCTC C	Reverse primer to confirm $nam7_{W\text{-}UM}$
6108-1	GTATGGCGGTCGTTTCAC CGTC	Forward primer to amplify *IDH1*
6108-2	CTGGCGATCCTTTCTGTCT TAGG	Reverse primer to amplify *IDH1*
6078-1	AGACGTCGCGCTATACAG TTCAATCTACCAAT	Forward primer to amplify portion of *TRA1*
4225-1	TTTCATATGGAAATCTTG AGGG	Reverse primer to amplify portion of *TRA1*

Table S3: DNA constructs used in this study

Name	Description
CB1503	*PHO5- LacZ*
pS14	*MET3* - Myc_9 - *NGG1* in YEplac195
pS17	*MET3* - Myc_9 - $TRA1_{PI3K/FATC}$ in YEplac195
pS26	*MET3* - Myc_9 - 6HIS-$TRA1_{PI3K/FATC}$ in YEplac195
CB1997	*MET3* - Myc_9 - *NGG1*
YEplac195	Multi-copy shuttle-vector plasmid containing *URA3* (Gietz, 1988)
CB1617	YCPlac33-*TRA1*

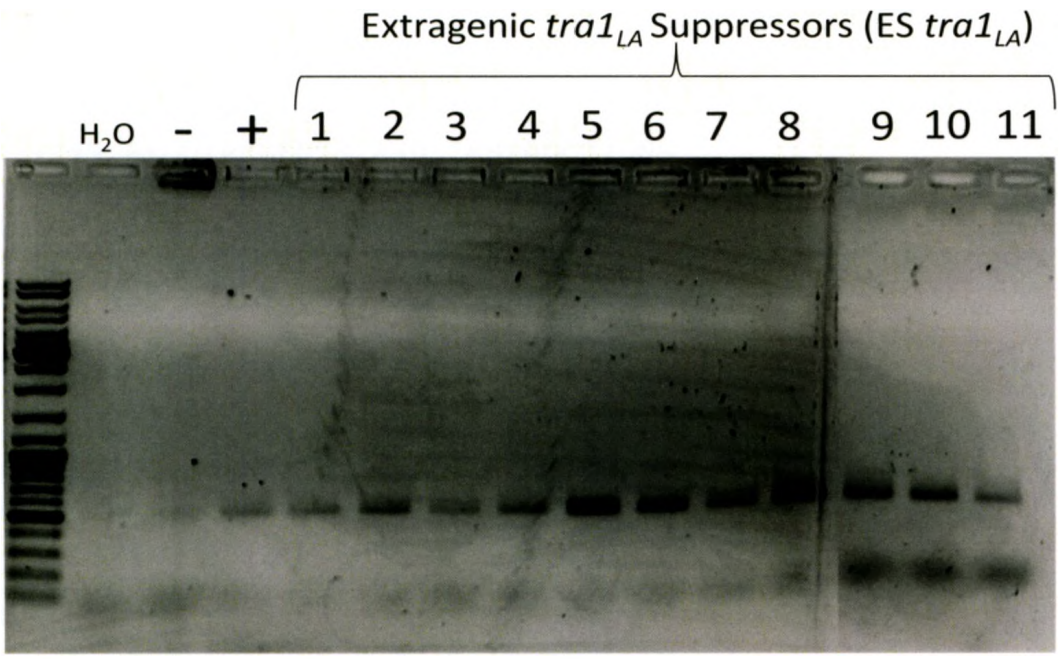

Figure S1: **PCR confirming presence of L3733A mutation in suppressor strains**. DNA from CY2706 (*TRA1*), CY4018 ($tra1_{LA}$), CY5579 (*es2* $tra1_{LA}$), CY5580 (*es12* $tra1_{LA}$), CY5581 (*es35* $tra1_{LA}$), CY5582 (*es36* $tra1_{LA}$), CY5583 (*es37* $tra1_{LA}$), CY5584 (*es38* $tra1_{LA}$), CY5585 (*es39* $tra1_{LA}$), CY5586 (*es40* $tra1_{LA}$), CY5587 (*es41* $tra1_{LA}$), CY5588 (*es42* $tra1_{LA}$), and CY5750 (*es43* $tra1_{LA}$), was amplified by PCR. Forward primer 4292-1 and reverse primer 5632-1 containing a mismatch to the leucine codon at nucleotide position 11197-11198 of *TRA1* and a match to alanine, were used to confirm the presence of a 512 base pair PCR fragment. PCR products were separated on a 1 % agarose gel electrophoresis and visualized under UV after ethidium bromide staining to check for the desired band. No fragment is seen for the water control (H_2O) and a very faint band is seen for the negative control containing DNA from *TRA1* (-). PCR fragments corresponding to 512 base pairs are seen for: the positive control $tra1_{LA}$ (+), and suppressors *es2* $tra1_{LA}$ (1), *es12* $tra1_{LA}$ (2), *es35* $tra1_{LA}$ (3), *es36* $tra1_{LA}$ (4), *es37* $tra1_{LA}$ (5), *es38* $tra1_{LA}$ (6), *es39* $tra1_{LA}$ (7), *es40* $tra1_{LA}$ (8), *es41* $tra1_{LA}$ (9), *es42* $tra1_{LA}$ (10), and *es43* $tra1_{LA}$ (11).

```
##this script will convert the files to BAM format, sort and
index them
#the variants are called from the bam file
#
###### REQUIREMENTS AND INSTRUCTIONS
#YOU MUST HAVE BOWTIE AND SAMTOOLS (0.1.11) INSTALLED AND IN YOUR
PATH
#YOU NEED A REFERENCE GENOME WHICH MUST BE THE SAME ONE THAT THE
INDEX WAS BUILT ON
#
##### REQUIRED VARIABLES
#
#location and names of the starting files
FASTA=solid0332_20110201_PE_BC_2_KH_CB_Pool_CB_F3_
FASTA1=solid0332_20110201_PE_BC_2_KH_CB_Pool_CB_F5-BC_
QUAL=solid0332_20110201_PE_BC_2_KH_CB_Pool_CB_F3_QV_
QUAL1=solid0332_20110201_PE_BC_2_KH_CB_Pool_CB_F5-BC_QV_
#
#base name for the BOWTIE index file
IDX=yeast
#
##these are the base names for each of the samples
## The samples are all named the same except for this one
difference
#Specify the SET
SET=(SK1 SKPSE2 SKPSE38 SKPSE41) #These are Stephanie's samples
SET=(CB1132 CB1134X CB7321) #These are Chris' samples
SET= (IS1 IS2 IS3 IS4 IS5 IS6) #These are Ivan's yeast samples

#get the number of elements in the SET array
NSET=(${#SET[@]})
#
#these are the chromosome names as they appear in the bowtie
output file
CHR=(NC_001133 NC_001134 NC_001135 NC_001136 NC_001137 NC_001138
NC_001139 NC_001140 NC_001141 NC_001142 NC_001143 NC_001144
NC_001145 NC_001146 NC_001147 NC_001148)
#and the number of chromosomes
NCHR=(${#CHR[@]})
#
########## PROGRAM STARTS HERE ##########################
#
########## MAKE the BOWTIE COLORSPACE INDEX
########## Do this once only so check to see if they already
exist
if [ -e $IDX.1.ebwt ]; then
echo "index files exist"
else
bowtie-build -f -C yeast.fsa $IDX
fi

#### run through each name in the set
```

```
for (( i=0; i<=$NSET; i++ ));
do

#base directory for this sample
DIR=(${SET}_forward_dir)

#this sample
SAMPLE=(${SET})

#set up the directory structure
#each bam file goes in its own directory named X_var_dir
     done
     if [ -d $DIR ]; then
         echo $DIR exists
else
         echo making $DIR
         mkdir $DIR
     fi
echo "running set $SAMPLE"


########## MAP the reads using bowtie
########## Do this once only
########## INPUT is CS reads, CS quality
########## OUTPUT is to map_best_SAMPLE.out
echo "mapping with bowtie"
$FASTA$SAMPLE.csfasta -f -C -2 $FASTA1$SAMPLE.csfasta --Q1
$QUAL$SAMPLE.qual --Q2 $QUAL1$SAMPLE.qual
$DIR/map_best1_time_$SAMPLE.out
bowtie -S -C -v 3 -t $IDX -f $FASTA$SAMPLE.csfasta -Q
$QUAL$SAMPLE.qual $DIR/map_best1_$SAMPLE.out
bowtie -S -C -3 10 --best -M 1 -v 3 -t $IDX -f
$FASTA1$SAMPLE.csfasta -Q $QUAL1$SAMPLE.qual
$DIR/map_best2_$SAMPLE.out
#Use -f to specify the fasta .csfasta files and -Q (for unpaired
reads) or --Q1/--Q2 (for paired-end reads)
#to specify the corresponding _QV.qual files.
#-S means to print in SAM format and -C means colourspace
#-3 for paired-end alignments (only uses ebwt.3 and ebwt.4)
#--best and -M 1 mean to report only the best alignment and not
more than one alignment
#wc -l to count number of reads
#-3 10 to take 10 nt off C-terminus

echo "beginning sorting - be patient!"
#extract out each chromosome for sorting
for (( j=0;j<$NCHR;j++ )); do

         #set the name
         CHROM=${CHR[$j]}
         echo "doing $CHROM $NCHR on
$DIR/map_best1_trim10_$SAMPLE.out"
```

```
        echo "saving to $DIR/$SAMPLE$CHROM.out"

            ########## EXTRACT OUT EACH CHROMOSOME OR FASTA FILE
SO THEY CAN BE SORTED INDIVIDUALLY
########## THIS IS MORE EFFICIENT THAN SORTING IN ONE BIG FILE
        awk -v CHR=$CHROM '$3 ~ CHR' $DIR/map_best1_$SAMPLE.out
> $DIR/$SAMPLE$CHROM.out
#The 3rd column contains the NC number. Use ~ because its not an
exact match

########## SORT READS IN EACH FILE SEQUENTIALLY, 4, 4n means sort
column 4 numerically
        sort -k 4,4n $DIR/$SAMPLE$CHROM.out >
$DIR/sorted_$SAMPLE$CHROM.out

########## MERGE THE SORTED FILES by appending
        cat $DIR/sorted_$SAMPLE$CHROM.out >>
$DIR/sorted_$SAMPLE.out

        rm $DIR/sorted_$SAMPLE$CHROM.out
        rm $DIR/$SAMPLE$CHROM.out
#done
echo "done sorting"
#
#add the header information to the sorted sam files, need first
18 lines
head -n 18 $DIR/map_best1_$SAMPLE.out > $DIR/head_$SAMPLE.txt
echo "making $DIR/sorted_$SAMPLE.sam"
cat  $DIR/head_$SAMPLE.txt $DIR/sorted_$SAMPLE.out >
$DIR/sorted_$SAMPLE.sam

echo "converting to bam"

#convert files to BAM format
samtools view -bS -o $DIR/sorted_$SAMPLE.bam
$DIR/sorted_$SAMPLE.sam

#The header information is needed for this step
samtools view -bS $DIR/sorted_$SAMPLE.sam >
$DIR/sorted_$SAMPLE.bam

#The -b means output in bam format, -S means input is in SAM, -o
means output file - ie: bam

echo "sorting bam file"
#sort the BAM file
samtools sort $DIR/sorted_$SAMPLE.bam $DIR/bamsorted_$SAMPLE
samtools sort <in> <out>
#
echo "indexing bam file"

#index the BAM file
```

```
samtools index $DIR/bamsorted_$SAMPLE.bam

## "running pileup"
samtools pileup -vcf yeast.fsa $DIR/bamsorted_$SAMPLE.bam >
$DIR/raw_pileup_$SAMPLE.txt
samtools.pl varFilter $DIR/raw_pileup_$SAMPLE.txt | awk '$6>=20'
> $DIR/filtered_pileup_$SAMPLE.txt
#Phred quality score above 20 are filtered out
#VCF is variant call format
#
echo "cleaning up"
rm  $DIR/head_$SAMPLE.txt
rm $DIR/sorted_$SAMPLE.out
done
```

Figure S2: **Shell Script to map colourspace reads to the SGD yeast genome.** Bowtie was used to map reads to yeast genome, and a list of filtered polymorphisms with Phred score above 20 was obtained (sam_variants.sh). This script was written by Dr. Greg Gloor, and modified by Stephanie Kvas.

```
#!/usr/bin/env perl -w
use strict; #enforces good programing practices
use warnings;

=com
This program literally compares FILE2 to FILE1 and makes a third
file containing the differences in FILE2 not found in FILE1.
This program will be used to compare the polymorphisms between
wild-type whole-genome sequencing data and the suppressors.
Therefore, this will be used to narrow down the number of
candidate polymorphisms.
=cut

my $filename = $ARGV[0]; #specify first file to compare to
my $filename2 = $ARGV[1]; #specify second file to compare to
my $outfile = $filename . "_final.txt"; #Thefinal file
my %results = ();

open FILE1, "$filename" or die "$!\n";
while(my $line = <FILE1>){
$results{$line} = 1;
}
close(FILE1);


open FILE2, "$filename2" or die "$!\n";
while(my $line =<FILE2>) {
$results{$line}++;
}
close(FILE2);

open (OUTFILE, ">$outfile") or die "$!\n";
foreach my $line (keys %results) {
print OUTFILE $line if $results{$line} == 2;
}
close OUTFILE;
```

Figure S3: **Perl script to eliminate background polymorphisms**. Background mutations found in wild-type samples were eliminated (Comparisons.pl). This script was written by Stephanie Kvas.

```
#!/usr/bin/env perl -w
use strict;

#read file
my $rfile = $ARGV[0];
#feature table
my $ftfile = $ARGV[1];

#header for the output table, this is the set
my $header = $ARGV[2];
#the number of iterations
my $r = $ARGV[3];

#hash of array of reads, keyed by node
my %freads;
my %rreads;
my %wt;
my %mut;

###Pull out the information of interest to me####
open (IN, "< $rfile") or die "$!\n";
        while(defined(my $l = <IN>)){
                chomp $l;
                my @l = split/\t/, $l;
#               ${ $freads{$l[0]} }[$l[1]] = $l[2];
#               ${ $rreads{$l[0]} }[$l[1]] = $l[3];
#               ${ $reads{$l[0]} }[$l[1]] = $l[3];
                ${ $wt {$l[0]} } = $l[2];
                ${ $mut {$l[0]} } = $l[3];
        }
close IN;

my @data;
#run through all feature tables in the set
#output is by redirection
        open (IN, "< $ftfile") or die "$!\n";
                while(defined(my $l = <IN>)){
                chomp $l;
                        push @data, $l;
                }
        close IN;

for(my $i = 0; $i < @data; $i++){
        my $l = $data[$i];
        my $ln = $data[$i+1];

        #print "$l\t$ARGV[2]_fwdr\t$ARGV[2]_revr\n" if ($l =~
/^#/);

        if ($l !~ /^#/){
                my @l = split/\t/, $l;
                $l[2] =~ s/\D//; $l[3] =~ s/\D//;
```

```
                my $max = $l[2]; my $min = $l[3];
                $max = $l[3] if $l[3] > $max;
                $min = $l[2] if $l[2] < $min;
                my $fr = my $rr = 0;
                for(my $i = $min; $i <= $max; $i++){
                        $fr += ${ $freads{$l[1]} }[$i] if exists
${ $freads{$l[1
]} }[$i];
                        $rr += ${ $rreads{$l[1]} }[$i] if exists
${ $rreads{$l[1
]} }[$i];
                }
                #print "$l[0]\t$l[-
1]\t$l[1]\t$l[2]\t$l[3]\t$fr\t$rr\n" if $r ==
 0; #works
                #print "$l\t$fr\t$rr\n" if $r > 0; #Does not work

                #now find intergenic reads
                my @ln = split/\t/, $ln;
                my $maxln = $l[2]; my $minln = $l[3];
                $maxln = $l[3] if $l[3] > $maxln;
                $minln = $l[2] if $l[2] < $minln;

                my $frln = my $rrln = 0;
                for(my $i = $max; $i < $minln; $i++){
                        $frln += ${ $freads{$l[1]} }[$i] if
exists ${ $freads{$l
[1]} }[$i];
                        $rrln += ${ $rreads{$l[1]} }[$i] if
exists ${ $rreads{$l
[1]} }[$i];
                }
                #print "$l[0]\t$l[-
1]\t$l[1]\t$l[2]\t$l[3]\t$frln\t$rrln\n" if $
r == 0;
                #print "$l\t$frln\t$rrln\n" if $r > 0;

        }

}
```

Figure S4: **Perl script to assign gene names and descriptions to each polymorphism** (add_reads_to_features.pl). This script was written by Dr. Greg Gloor with slight modifications made by Stephanie Kvas.