

1-1-2014

# Genetic and Proteomic Analysis of Rna Polymerase Ii-Interacting Complexes Mediator and Integrator in the Ciliated Protozoan Tetrahymena Thermophila

Matthew D.R. Cadorin  
*Ryerson University*

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>



Part of the [Molecular genetics Commons](#)

---

## Recommended Citation

Cadorin, Matthew D.R., "Genetic and Proteomic Analysis of Rna Polymerase Ii-Interacting Complexes Mediator and Integrator in the Ciliated Protozoan Tetrahymena Thermophila" (2014). *Theses and dissertations*. Paper 2015.

GENETIC AND PROTEOMIC ANALYSIS OF RNA POLYMERASE II-INTERACTING  
COMPLEXES MEDIATOR AND INTEGRATOR IN THE CILIATED PROTOZOAN

*TETRAHYMENA THERMOPHILA*

by

Matthew D.R. Cadorin

B.Sc., Biology, McMaster University, 2009

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Science

in the program of

Molecular Science

Graduate program in Chemistry and Biology

Ryerson University

Toronto, Ontario, Canada

© Matthew D.R. Cadorin, 2014

I hereby declare that I am the sole author of this thesis or dissertation.

I authorize Ryerson University to lend this thesis or dissertation to other institutions or individuals for the purpose of scholarly research.

Matthew D.R. Cadorin

I further authorize Ryerson University to reproduce this thesis or dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Matthew D.R. Cadorin

**GENETIC AND PROTEOMIC ANALYSIS OF RNA POLYMERASE II-INTERACTING  
COMPLEXES MEDIATOR AND INTEGRATOR IN THE CILIATED PROTOZOAN  
*TETRAHYMENA THERMOPHILA*.**

Matthew D.R. Cadorin, Master of Science, Molecular Science, Ryerson University, 2014

**Abstract**

In most eukaryotes, the largest subunit of RNAPII, Rpb1, contains a conserved carboxy-terminal domain (CTD) containing a canonical structure of heptapeptide repeats. Two protein complexes of interest, Mediator and Integrator, are known to interact with this CTD in all eukaryotic models they have been described in to date. Recently, orthologs of Mediator and Integrator subunits have been identified within the ciliated protozoan *Tetrahymena thermophila*; one of the few eukaryotic lineages to lack a canonically organized CTD.

To begin to characterize putative Mediator and Integrator complexes within *T. thermophila*, I engineered appropriate macronuclear tagging and knockout cassettes. Although the *Tetrahymena MED31* ortholog was unable to rescue the slow growth phenotype of a yeast *MED31* knockout, or co-purify with yeast Med8-TAP, I identified subunit Med3 as a member of the Med31 interactome in *T. thermophila* through tandem affinity purification coupled with mass spectrometry. I also targeted the *Tetrahymena INTS6* locus for knockout as determined by colony PCR. If Mediator and Integrator exist in *Tetrahymena* despite its divergent CTD of Rpb1, perhaps these complexes have CTD-independent functions beyond what can be effectively studied using conventional model systems.

## **Acknowledgements**

I express my great appreciation toward Dr. Jeffrey Fillingham for providing the wonderful opportunity to learn and grow under his supportive and stimulating guidance. I am also particularly grateful for the generous assistance and knowledgeable insight provided by Dr. Jyoti Garg throughout my research. I must also express my gratitude toward Dr. Ron Pearlman at York University for the equipment and facilities kindly made accessible by him, and to Dr. Jean Philippe Lambert at the Lunenfeld-Tanenbaum Research Institute for his expertise at mass spectrometry and eagerness to serve our needs. I would also like to recognize the contributions of the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding this pursuit of knowledge and understanding.

This journey would not have been as enjoyable without the mutual support, advice, and laughter shared with colleagues, particularly Syed Nabeel Hyder Shah, Ernest Radovani, Kanwal Ashraf, and Nora Saud Dannah. My gratitude extends to the staff of the Department of Chemistry and Biology of Ryerson University for their selfless efforts behind the scenes.

I owe an immeasurable debt to my parents, Virginia and Derek, and my sister Megan for their unwavering and unconditional support, and for ensuring I always got to the bus on time. Lastly, I thank my friends and family for keeping me grounded through the unpredictable nature of science, and for their patience and encouragement throughout my endeavour, especially during the inevitable sacrifices it sometimes entailed.

# Table of Contents

Title Page .....	i
Author's declaration .....	ii
Abstract .....	iii
Acknowledgements .....	iv
List of Tables .....	ix
List of Figures .....	x
List of Appendices .....	xii
List of Abbreviations .....	xiii
Chapter 1: Introduction .....	1
1.1 RNA polymerase II .....	1
1.1.1 RpbI .....	1
1.2 Carboxy-terminal domain (CTD) .....	2
1.2.1 Phosphorylation code .....	3
1.2.1.1 Serine <sup>2</sup> and serine <sup>5</sup> phosphorylation .....	3
1.2.1.2 Serine <sup>7</sup> phosphorylation .....	4
1.2.1.3 Serine phosphatases .....	5
1.3 Rpb1 CTD conservation .....	6
1.4 <i>Tetrahymena thermophila</i> .....	7
1.4.1 Nuclear dimorphism in <i>Tetrahymena thermophila</i> .....	8
1.4.2 Vegetative life cycle of <i>Tetrahymena thermophila</i> .....	8
1.4.3 Sexual reproduction in <i>Tetrahymena thermophila</i> .....	10
1.5 Programmed DNA rearrangements .....	11
1.5.1 Breakage eliminated sites (BESs) .....	12
1.5.2 Internal eliminated sequences (IESs) .....	12
1.5.2.1 Scan RNA model .....	14
1.6 <i>Tetrahymena thermophila</i> : challenges and considerations .....	15
1.6.1 Highly polyploid macronucleus .....	15
1.6.2 AT-rich genome .....	15

1.6.3 Codon usage in <i>Tetrahymena</i> .....	16
1.7 Studying CTD-interacting proteins in <i>Tetrahymena thermophila</i> .....	17
1.8 Mediator complex .....	17
1.8.1 Structure and function of Mediator.....	18
1.8.2 Mediator conservation .....	20
1.8.3 Identification of <i>Tetrahymena</i> Mediator.....	21
1.9 Integrator complex .....	22
1.9.1 Structure and function of Integrator.....	23
1.9.2 Identification of <i>Tetrahymena</i> Integrator.....	24
1.10 Tandem mass spectrometry and SAINT .....	25
1.11 Rationale and project summary .....	27
Chapter 2: Materials and Methods .....	29
2.1 Equipment.....	29
2.2 Sequence alignments.....	29
2.3 Growth conditions.....	30
2.4 Media, buffers, solutions.....	31
2.5 Manual <i>T. thermophila</i> genomic DNA extraction .....	31
2.6 Manual <i>E. coli</i> plasmid DNA miniprep (alkaline lysis miniprep).....	32
2.7 Polymerase chain reaction (PCR) .....	33
2.8 DNA restriction digest/linearization .....	33
2.9 Enzymatic cleanup and gel extraction .....	34
2.10 DNA electrophoresis.....	34
2.11 DNA ligation and transformation into competent <i>E. coli</i> .....	34
2.12 Sequencing.....	35
2.12.1 <i>E. coli</i> plasmid DNA isolation for sequencing .....	35
2.13 DNA purifications.....	36
2.13.1 PEG purification .....	36
2.13.2 Ethanol purification .....	36
2.14 Construction of the 3xFLAG-TEV-ZZ (FZZ) tagging cassette.....	37
2.15 Construction of the knockout cassette .....	38
2.16 Biolistic transformation of <i>T. thermophila</i> .....	39

2.16.1 Preparation of <i>T. thermophila</i> cells.....	39
2.16.2 Preparation of gold beads.....	40
2.16.3 Preparation of flying discs .....	40
2.16.4 Assembly and operation.....	40
2.16.5 Selection.....	41
2.17 Rapid <i>T. thermophila</i> DNA extraction/colony PCR.....	41
2.18 Preparation of <i>T. thermophila</i> cell extracts (TCA extraction) .....	42
2.19 Western blot analysis .....	42
2.19.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) .....	42
2.19.2 Western transfer .....	43
2.19.3 Ponceau stain .....	44
2.19.4 Blocking.....	44
2.19.5 Probing.....	44
2.20 Tandem affinity purification (TAP) in <i>T. thermophila</i> .....	45
2.20.1 Growing large cultures of <i>T. thermophila</i> .....	45
2.20.2 Preparation and clarification of <i>T. thermophila</i> whole cell extracts (WCEs).....	45
2.20.3 Preparation of IgG-Sepharose.....	46
2.20.4 TEV cleavage.....	47
2.20.5 Preparation of M2-agarose.....	48
2.20.6 Final elution .....	48
2.21 Mass spectrometry .....	48
2.22 Transformation of <i>S. cerevisiae</i> .....	50
2.23 Engineering of yeast and <i>Tetrahymena MED31</i> expression vectors .....	51
2.24 Affinity purification in <i>S. cerevisiae</i> .....	52
Chapter 3: Results .....	54
3.1 RNA polymerase II largest subunit Rpb1 lacks a canonical CTD in <i>Tetrahymena</i> .....	54
3.2 Mediator subunit Med31 is conserved in eukaryotes .....	57
3.3 <i>MED31</i> of <i>T. thermophila</i> does not rescue the slow-growth phenotype of <i>S. cerevisiae</i> deletd for <i>MED31</i> .....	58
3.4 12myc-Med31 <sup>Tt</sup> does not co-purify with Med8 <sup>Sc</sup> -TAP .....	61
3.5 Construction of 3xFLAG-TEV-ZZ tagging cassette .....	63

3.5.1 PCR amplification of 5' and 3' homology sequences.....	68
3.5.2 Diagnostic restriction digest .....	70
3.5.3 Confirmation of FZZ-tagging cassettes .....	70
3.5.4 Linearization of tagging construct for transformation into <i>T. thermophila</i> .....	71
3.6 Int4-FZZ and Med20-FZZ screening Western blots.....	72
3.7 Confirmation of FZZ tagging of <i>Tetrahymena</i> Med22.....	73
3.8 Med22 affinity purification-mass spectrometry in <i>Tetrahymena</i> .....	74
3.9 Med31 affinity purification-mass spectrometry in <i>Tetrahymena</i> .....	76
3.10 Int11-FZZ affinity purification-mass spectrometry in <i>Tetrahymena</i> .....	79
3.11 Construction of the <i>INTS6</i> knockout cassette .....	80
3.11.1 Successful knockout of <i>INTS6</i> (colony PCR).....	83
Chapter 4: Discussion .....	90
4.1 Rpb1 of <i>Tetrahymena</i> RNA polymerase II lacks a canonical CTD.....	90
4.2 Med22 AP-MS .....	92
4.3 Med31 AP-MS .....	93
4.4 Int11 AP-MS.....	94
4.5 Speculative role of <i>Tetrahymena</i> Med31 in meiotic transcription.....	96
4.6 Tagging <i>INTS6</i> with a C-terminal FZZ epitope tag in <i>T. thermophila</i> .....	98
4.7 Future prospective of Integrator knockout.....	99
Conclusion .....	100
Appendix A: Fms1 and Vps75.....	102
Appendix B: Rtt109 .....	106
Appendix C: Supplementary Material .....	111
References.....	128

## List of Tables

Table 1. PCR conditions for 2x PrimeSTAR Max DNA Polymerase (TaKaRa) .....	33
Table 2. Percent abundance of serine, threonine, tyrosine, and proline in various Rpb1 CTDs ..	57
Table 3. Mass spectrometry data for <i>Tetrahymena</i> Dss1-FZZ .....	66
Table 4. Tandem mass spectrometry data for <i>Tetrahymena</i> Med22-FZZ .....	75
Table 5. SAINT analysis of <i>Tetrahymena</i> Med31-FZZ affinity purification .....	78

## List of Figures

Figure 1. Schematic of dynamic CTD phosphorylation .....	6
Figure 2. Schematic of <i>Tetrahymena</i> nuclei progression through its different life cycles. ....	9
Figure 3. Schematic of accelerated phenotypic assortment in <i>Tetrahymena</i> .....	11
Figure 4. Programmed DNA rearrangements in <i>Tetrahymena</i> .....	12
Figure 5. Interaction map of yeast Mediator complex .....	20
Figure 6. Multiple sequence alignment of the CTD of Rpb1 .....	56
Figure 7. Multiple sequence alignment of Med31 .....	58
Figure 8. Growth of <i>med31</i> <sup>Sc</sup> Δ yeast transformed with yeast and <i>Tetrahymena</i> <i>MED31</i> .....	60
Figure 9. Western blot of Med8-TAP affinity purification.....	63
Figure 10. Schematic of <i>Tetrahymena</i> FZZ-tagging vector pBKS-FZZ .....	67
Figure 11. Schematic of protein-tagging strategy in <i>Tetrahymena</i> .....	68
Figure 12. PCR products of 5' and 3' Mediator subunit homology sequences .....	69
Figure 13. Diagnostic restriction digest of representative FZZ tagging vectors.....	70
Figure 14. Gel electrophoresis of representative linearized tagging construct.....	71
Figure 15. Western blot of Med20-FZZ and Int4-FZZ screening .....	73
Figure 16. Western blot analysis of TCA-extracted putative <i>Tetrahymena</i> Med22-FZZ.....	74
Figure 17. Western blot analysis of putative <i>Tetrahymena</i> Med22 affinity purification.....	75
Figure 18. Western blot analysis of <i>Tetrahymena</i> Med31 affinity purification .....	77
Figure 19. Western blot analysis of <i>Tetrahymena</i> Int11 affinity purification .....	79
Figure 20. Schematic of <i>Tetrahymena</i> knockout vector p4T2-1 .....	81
Figure 21. Schematic of <i>INTS6</i> knockout strategy in <i>Tetrahymena</i> .....	82
Figure 22. <i>INTS6</i> KO diagnostic digests.....	83

Figure 23. Schematic of expected colony PCR product sizes for testing presence of *INTS6KO* cassette ..... 86

Figure 24. *INTS6KO* colony PCR with KpnIF/SacIIR primer pair ..... 87

Figure 25. Schematic of expected colony PCR primer binding for testing correct *INTS6KO* integration ..... 88

Figure 26. *INTS6KO* colony PCR with KpnIF/H4neoR and UF/H4neoR primer pairs ..... 89

## List of Appendices

Appendix A: Fms1 and Vps75.....	102
Appendix B: Rtt109 .....	106
Appendix C: Supplementary Material .....	111

## List of Abbreviations

Å	angstrom ( $1.0 \times 10^{-10}$ meters)
AP	affinity purification
APS	ammonium persulfate
BES	breakage eliminated site
BLAST	Basic Local Alignment Search Tool
bp	base pair
C-	carboxy
CBS	chromosome breakage site
CDK	cyclin-dependent kinase
CTD	carboxy-terminal domain
ddH <sub>2</sub> O	double distilled water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
FZZ	3x FLAG-ZZ epitope tag
GFP	green fluorescent protein
GST	glutathione S-transferase
HPLC	high-performance liquid chromatography
IES	internal eliminated sequence
IgG	immunoglobulin G

kb	kilo base pair
kDa	kilodalton
KO	knockout
LTQ	linear trap quadrupole
M	molar
MAC	macronucleus
MDS	macronuclear destined sequences
MIC	micronucleus
MP	minipreparation (DNA)
mRNA	messenger RNA
MS/MS	tandem mass spectrometry
μ	micro
MW	molecular weight
N-	amino
NP-40	nonidet P-40
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethanesulfonylfluoride
PSF	penicillin streptomycin fungizone
psi	pounds per square inch

RNA	ribonucleic acid
RNAi	RNA interference
rpm	rotations per minute
SAINT	significance analysis of interactome
scnRNA	scan RNA
SDS	sodium dodecyl sulfate
snRNA	small nuclear RNA
SPP	sequestrin protease peptone
SSM	signature sequence motif
TAP	tandem affinity purification
TBE	tris/borate/EDTA
TEMED	tetramethylethylenediamine
TEV	tobacco etch virus
UTR	untranslated region
UV	ultraviolet
v/v	volume/volume
WCE	whole-cell extract
WT	wildtype
w/v	weight/volume
YNB	yeast nitrogen base
YPD	yeast extract peptone dextrose
YT	yeast extract tryptone

## **Chapter 1: Introduction**

### **1.1 RNA polymerase II**

RNA polymerase (RNAP) is a DNA-dependent RNA polymerase that catalyzes the synthesis of RNA using DNA as a template in a process known as transcription. Five RNA polymerases exist in eukaryotes, each responsible for producing a unique subset of RNA products. RNAPI synthesizes precursors of large ribosomal RNA (pre-rRNA) involved in ribosome production (Russel and Zomwedijk, 2006). RNA polymerase II (RNAPII) is required for DNA transcription to produce the precursor messenger RNA (pre-mRNA) of all protein-coding genes, but also synthesizes small nuclear RNA (snRNA) and some micro RNA (miRNA) (Young, 1991; Lee *et al.*, 2004; Egloff *et al.*, 2008). In addition to its polymerase capacity, RNAPII also contains an associated exonuclease proofreading activity required for ensuring a high-fidelity product (Sydow and Cramer, 2009). RNAPIII is involved in the synthesis of a repertoire of non-coding RNAs including small nucleolar RNA (snoRNA), micro RNA, and transfer RNA (tRNA) (Dieci *et al.*, 2007). RNAPIV and RNAPV are plant-specific, and while their precise mechanics are still enigmatic, they are known to synthesize transcripts involved in the small interfering RNA (siRNA)-dependent formation of condensed, transcriptionally silenced DNA known as heterochromatin. (Wierzbicki *et al.*, 2008).

#### **1.1.1 RpbI**

RNAPII is a protein complex composed of 12 subunits (Rpb1-Rpb12) in yeast and humans (Young, 1991; Armache *et al.*, 2003). The largest subunit of RNAPII, Rpb1, is well

conserved across all domains of life, sharing eight homology regions (A-H) within eukaryotes, as well as with the  $\beta'$  subunit of RNAP in bacteria and the A' subunit of RNAP in Archaea (Allison, *et al.*, 1985; Sidow *et al.*, 1994). Following the last conserved homology region of Rpb1 (H domain) lies an amino acid extension termed the carboxy-terminal domain (CTD) connected to Rpb1 via a linker region (Matheny *et al.*, 2002; Appendix C.1). The Rpb1 CTD plays important roles in all facets of transcriptional regulation from transcription initiation (Kim *et al.*, 1994) through post-transcriptional 3' cleavage and polyadenylation (McCracken *et al.*, 1997). It also functions in alternative splicing (de la Mata and Kornblihtt, 2006) and the expression of small nuclear RNA (snRNA) (Medlin and Uguen, 2003).

## **1.2 Carboxy-terminal domain (CTD)**

Transcription, the process of replicating a segment of DNA into an RNA transcript, is an elaborate and highly synchronized process involving the coordinated efforts of many interacting proteins. In most eukaryotes, the CTD of the Rpb1 subunit of RNAPII contains a tandem array of Tyrosine<sup>1</sup>-Serine<sup>2</sup>-Proline<sup>3</sup>-Threonine<sup>4</sup>-Serine<sup>5</sup>-Proline<sup>6</sup>-Serine<sup>7</sup> (Y<sup>1</sup>S<sup>2</sup>P<sup>3</sup>T<sup>4</sup>S<sup>5</sup>P<sup>6</sup>S<sup>7</sup>) heptapeptide repeats, which are subject to extensive post-translational modification and serves as a scaffold for the systematic recruitment of transcription factors to RNAPII and the developing transcript (Allison *et al.*, 1988; Phatnani and Greenleaf, 2006). The Rpb1 CTD contains 26 tandem repeats in yeast and 52 in humans (Allison *et al.*, 1988), while other organisms such as *Tetrahymena thermophila* lack a definitive consensus sequence altogether (Stump and Ostrozhynska, 2013; Appendix C.2). The unique repeat structure of the Rpb1 CTD enables isomerization of the peptidyl-proline bonds (Wilcox *et al.*, 2004), glycosylation (Kelly *et al.*, 1993), ubiquitinylation (Li *et al.*, 2007), and methylation (Sims *et al.*, 2011), but it is the phosphorylation status of the

three serines, Ser<sup>2</sup>, Ser<sup>5</sup>, and Ser<sup>7</sup>, that mediates the sequential recruitment of specific nuclear factors (Bartkowiak *et al.*, 2011).

### **1.2.1 Phosphorylation code**

Phosphorylation is achieved by a class of enzymes known as protein kinases, which transfer phosphate groups to amino acid substrates in a process central to a wide variety of cellular events (Taylor and Kornev, 2011). Alternately, dephosphorylation is carried out in a reverse process by phosphatases (Mumby and Walter, 1993). The high substrate specificity of protein kinases enables targeting of particular residues within a polypeptide based on their relative position, such that serine kinases can distinguish between Ser<sup>2</sup>, Ser<sup>5</sup>, and Ser<sup>7</sup> of the CTD heptapeptides of Rpb1. Phosphorylation is differentially distributed across the serines in characteristic patterns corresponding to the position of RNAPII along the gene being transcribed. This phosphorylation code relates to particular events in the transcription cycle and provides a crucial mechanism for the temporal recruitment and displacement of CTD-interacting proteins lending a selective plasticity to the CTD (Komarnitsky *et al.*, 2000).

#### **1.2.1.1 Serine<sup>2</sup> and serine<sup>5</sup> phosphorylation**

Generally, phosphorylation of the Rpb1 CTD on Ser<sup>5</sup> (Ser<sup>5</sup>P) is abundant near promoter regions of genes, while that of Ser<sup>2</sup>P is only seen in coding regions (Komarnitsky *et al.*, 2000). Preceding transcription, RNAPII is hypophosphorylated and recruited to the promoter region of a gene during formation of the preinitiation complex (PIC) (Laybourn and Dahmus, 1989). CTD phosphorylation at this time is a negative regulator of PIC arrangement (Laybourn and Dahmus,

1990). Following the assembly of the transcription apparatus, phosphorylation of Ser<sup>5</sup> mediated by the cyclin-dependent kinase 7 (CDK7) catalytic subunit of transcription factor IIH (TFIIH) signals for the recruitment of mRNA capping enzymes and facilitates transcription initiation through release of promoter-proximal pausing (Cho *et al.*, 1997; Rodriguez *et al.*, 2000). In comparison, Ser<sup>2</sup> phosphorylation by the CDK9 catalytic subunit of the positive transcription elongation factor b (p-TEFb) grows more abundant as RNAPII nears the 3' end of genes, and marks the elongation and termination stages (Komarnitsky *et al.*, 2000; Cho *et al.*, 2001).

Therefore, as RNAPII moves along DNA during transcription from initiation, through elongation, and into termination, the phosphorylation status shifts from Ser<sup>5</sup>P to Ser<sup>2</sup>P (Figure 1). Phosphorylation of both Ser<sup>2</sup> and Ser<sup>5</sup> can also be catalyzed by Cyclin C/CDK8 (Sun *et al.*, 1998; Hsin and Manley, 2012), a member of the Mediator complex, although this action acts as a negative regulator of transcription by repressing TFIIH-mediated stimulation through CDK7 phosphorylation (Akoulitchev *et al.*, 2000).

#### **1.2.1.2 Serine<sup>7</sup> phosphorylation**

The role of Serine<sup>7</sup> phosphorylation has been more recently analyzed, and while both CDK7 and CDK9 have been shown to have Ser<sup>7</sup> kinase activity (Glover-Cutter *et al.*, 2009), its functional importance is still largely enigmatic (Figure 1). Despite being present during transcription of all RNAPII-transcribed genes, Ser<sup>7</sup>P does not appear to be essential for the transcription of protein-coding genes, as its substitution for unphosphorylatable alanine does not affect mRNA expression (Egloff *et al.*, 2007). The phosphorylation of Ser<sup>7</sup> is, however, essential for facilitating the expression of a class of non-coding RNAs known as small nuclear RNA (snRNA); a major component of the spliceosome (Egloff *et al.*, 2007). Unlike their protein-

coding counterparts, snRNA-encoding genes are neither spliced nor polyadenylated, but rather contain a conserved RNA-processing element downstream of the gene in the form of a 3' box, which is cleaved by the Integrator protein complex (Hernandez, 1985; Baillat *et al.*, 2005). Although the CTD of Rpb1 had been previously implicated in snRNA 3' processing, it was discovered that Ser<sup>7</sup>P in conjunction with Ser<sup>2</sup>P in a unique Ser<sup>2</sup>/Ser<sup>7</sup> double phosphorylation mark across adjacent heptapeptide repeats was a prerequisite specifically recognized by Integrator for its binding to the CTD (Uguen *et al.*, 2003; Egloff *et al.*, 2010).

### **1.2.1.3 Serine phosphatases**

It is important to remember that it is not simply a matter of phosphorylation, but the dynamic dephosphorylation by serine phosphatases that contributes to the phosphorylation code of the CTD. Dephosphorylation is also necessary for the eventual recycling of RNAPII back to its hypophosphorylated state. To accomplish this, Fcp1 preferentially dephosphorylates Ser<sup>2</sup>P (Cho *et al.*, 2001), while yeast Ssu72 and Rtr1 preferentially dephosphorylate Ser<sup>5</sup>P and are essential for the transition from Ser<sup>5</sup>P to Ser<sup>2</sup>P (Krishnamurthy *et al.*, 2004; Mosley *et al.*, 2009; Figure 1).

The seemingly simple sequence of repeats in Rpb1 can therefore be viewed as an organizing center, which coordinates RNAPII-mediated transcription and effectively couples it with co-transcriptional RNA processing of mRNA and 3' end processing of snRNA.

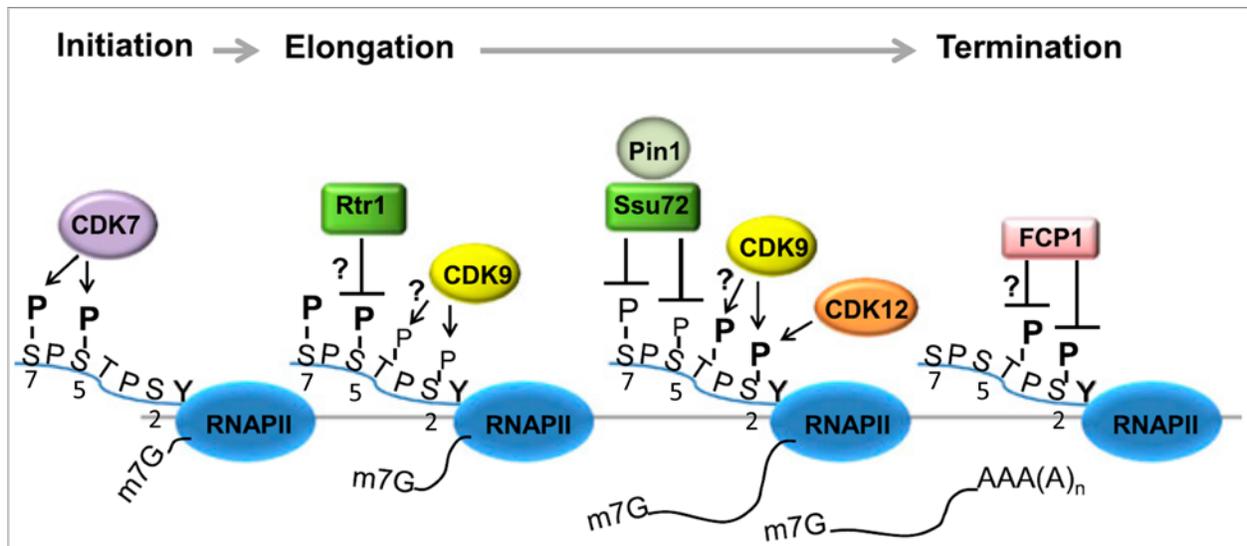


Figure 1. Schematic of dynamic CTD phosphorylation

A host of kinases and phosphatases modulate the dynamic phosphorylation status of the CTD to enable its selective recruitment of transcriptional proteins (adapted from Hsin and Manley, 2012).

### 1.3 Rpb1 CTD conservation

The CTD of Rpb1 is conserved in metazoans from fungi to humans, and the number of tandem repeats seems to be correlated with genetic complexity and varies between species, with the budding yeast *Saccharomyces cerevisiae* containing 26, fruit fly *Drosophila melanogaster* with 44, and vertebrates containing 52 consensus repeats (Allison *et al.*, 1988). Not all repeats strictly resemble the canonical Y<sup>1</sup>S<sup>2</sup>P<sup>3</sup>T<sup>4</sup>S<sup>5</sup>P<sup>6</sup>S<sup>7</sup> sequence, and copies nearing the 3' end of the CTD often deviate to contain amino acid substitutions (Hsin and Manley, 2012; Appendix C.3). Interestingly, serial CTD truncations in both yeast and mice determined that these divergent repeats were dispensable for viability, and cells were still viable with CTDs consisting of only 50% of their original heptapeptides (Bartolomei *et al.*, 1988; West and Corden, 1995). Below this threshold, cells are no longer viable *in vivo* most likely due to inadequate regulation of

mRNA transcription and RNA processing; processes mediated by the CTD of Rpb1. These same studies performed *in vitro*, however, revealed RNAPII still retains its ability to recognize promoter regions and initiate transcription (Zehring *et al.*, 1988). Studies in the protist *Trypanosoma brucei* have shown that even non-canonical CTDs of Rpb1 can be essential for viability (Das and Bellofatto, 2009), and so perhaps divergent systems lacking canonical CTDs on Rpb1, such as the ciliated protozoan *Tetrahymena thermophila*, contain an equally divergent means of RNA processing.

#### **1.4 Tetrahymena thermophila**

*Tetrahymena thermophila* is a single-celled, ciliated protozoan that lives freely in fresh water systems. It is a member of the phylum Ciliophora (Ciliates) together with other representative genera Paramecium, Oxytricha, and Ichthyophthirius. On a larger scale, the Ciliates, Dinoflagellates, Apicomplexa, and Chromerida phyla compose the Alveolates. The genus "*Tetrahymena*" is named after its four (tetra, Gr. = four) membrane-like (hymen, L. = membrane) oral structures: its primary undulating oral membrane and three oral "membranelles" of clustered cilia known as polykinetids (Lynn and Doerder, 2012). Ciliates diverged early in the eukaryotic lineage, before the establishment of metazoic fungi or animals (Baldauf, 2003).

Despite being distantly related to humans, there are 2,280 human genes with orthologs in *T. thermophila*, 874 of which are not found in the yeast *S. cerevisiae*. In fact, there are more orthologs shared between humans and *Tetrahymena* than there are between humans and yeast (Eisen *et al.*, 2006). As a defining feature of ciliates, *Tetrahymena* display many cilia that line the exterior of their cell membrane that are used as the basis for locomotion. Many studies have exploited this feature, but perhaps more remarkable is the involvement of *T. thermophila* in the

seminal discoveries of telomerase and catalytic RNA, both of which led to Nobel prizes (Kruger *et al.*, 1982; Greider and Blackburn, 1985). *Tetrahymena's* impact as an experimental system can be attributed in part to experimentally favourable features including large cell size, a rapid doubling time of two hours, nuclear dimorphism, genetic tractability, and a general ease of culturing (Collins, 2012).

#### **1.4.1 Nuclear dimorphism in *Tetrahymena thermophila***

Characteristic of ciliates, *Tetrahymena* utilize nuclear dimorphism as a strategy to separate the functionally and structurally differing germline and somatic nuclei within a single cytoplasm. The smaller germ line micronucleus (MIC) is diploid and contains 5 pairs of chromosomes that are coiled into transcriptionally silent heterochromatin. (Gorovsky and Woodard, 1969). In comparison, the larger somatic macronucleus (MAC) is transcriptionally active and is highly polyploid, with approximately 45 copies of each of its ~225 chromosomes (Eisen *et al.*, 2006).

#### **1.4.2 Vegetative life cycle of *Tetrahymena thermophila***

During the vegetative growth stage of *Tetrahymena*, the two nuclei replicate and divide independently of each other (Figure 2, stage V1). Only the micronucleus is capable of dividing by mitosis, and equally segregates to supply each of the two daughter micronuclei with identical genetic content (Ray, 1956; LaFountain and Davidson, 1980; Eisen *et al.* 2006). The chromosomes of the macronucleus are acentromeric (Cervantes *et al.*, 2006) and thus lack the attachment point for the centriolar spindle fibers required to pull apart the duplicated chromosomes to produce two genetically equal nuclei. Instead, the macronucleus undergoes

amitosis, a process reminiscent of binary fission whereby genetic material is randomly partitioned between the two daughter nuclei. This is the foundation for phenotypic assortment (Figure 3), and heterozygotes under selective pressure can become homozygous for a particular gene within 100 generations (Orias and Newby, 1975). Researchers can exploit this unique macronuclear behaviour through accelerated phenotypic assortment via drug resistance by the cell to experimentally generate cells homozygous for a particular gene (Merriam and Bruns, 1988).

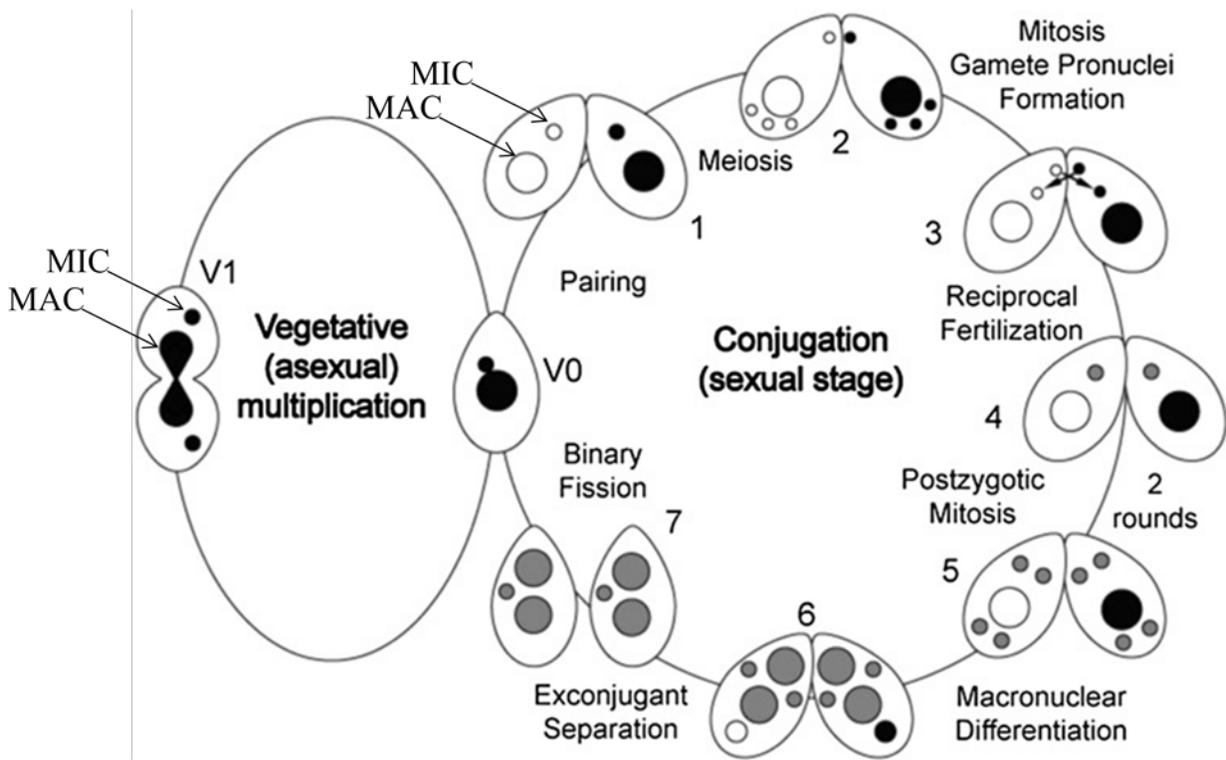


Figure 2. Schematic of *Tetrahymena* nuclei progression through its different life cycles.

*Tetrahymena* nuclei divide independently of each other during vegetative growth and sexual stages (adapted from Orias *et al.*, 2011).

### 1.4.3 Sexual reproduction in *Tetrahymena thermophila*

When food supply is insufficient to permit vegetative growth and cell division, *Tetrahymena* undergo sexual reproduction with a cell of complimentary mating type (Figure 2, stage 1). *Tetrahymena* generates seven different mating types (I to VII), and mating must take place between cells of different mating types (Elliott and Hayes, 1953; Cervantes *et al.*, 2013).

The two nuclei differ in their outcome of sexual reproduction. In preparation for conjugation, the micronucleus divides by meiosis to produce four haploid nuclei (Figure 2, stage 2). One of the four is chosen while the other three are ablated. This gametic micronucleus then undergoes one round of mitosis to duplicate itself into two nuclei. Allogamy occurs next when mating cells exchange one of the two haploid pronuclei while the other remains within the parental cell (Figure 2, stage 3). Once migration of the nuclei has taken place, the parental and the inherited pronuclei fuse to form a zygotic product (Figure 2, stage 4). This diploid nucleus of micronuclear origin divides twice by mitosis to form four daughter diploid nuclei (Figure 2, stage 5); two daughter micronuclei and two which develop into macronuclei (figure 2, stage 6) through the elimination of germ line-specific DNA (Karrer *et al.*, 1984; Karrer, 2012). The parental MAC is still present while the daughter MACs are developing, but it soon degenerates by what is proposed to be lysosomal autophagy (Figure 2, stage 7) (Akematsu *et al.*, 2010). Macronuclear gene expression therefore switches from the parental macronucleus, which coordinates early conjugation, to the zygotic macronucleus, which assumes the responsibility in late conjugation. The vegetative nuclear organization is restored following the first post-conjugation cell division (Figure 2, stage V1).

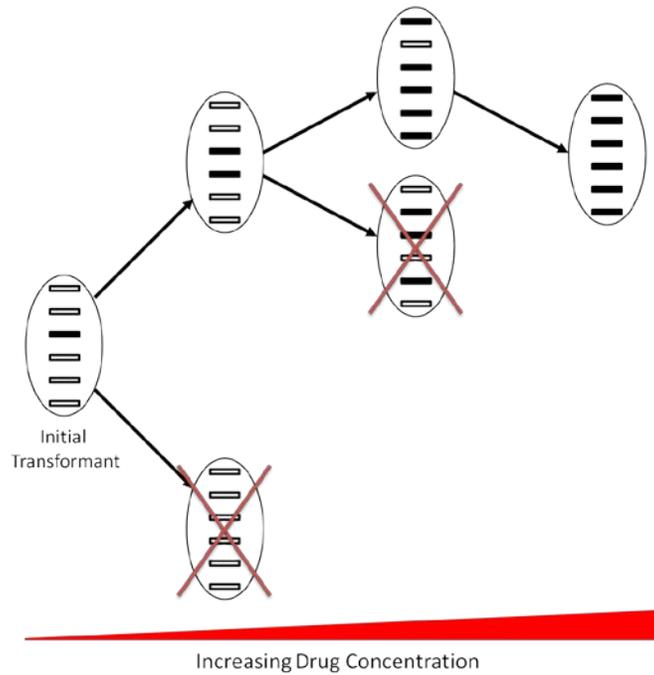


Figure 3. Schematic of accelerated phenotypic assortment in *Tetrahymena*

The macronucleus divides by amitosis resulting in unequal segregation of chromosomes. Selective pressure drives increasing copies of particular alleles within each generation until homozygosity is achieved.

### 1.5 Programmed DNA rearrangements

During later stages of conjugation, the developing somatic macronucleus undergoes extensive programmed DNA rearrangements, which accounts for elimination of around 15% of the macronuclear genome (Yao *et al.*, 1984). The programmed DNA rearrangements consists of breakage eliminated sites (BESs) and internal eliminated sequences (IESs) (Figure 4).

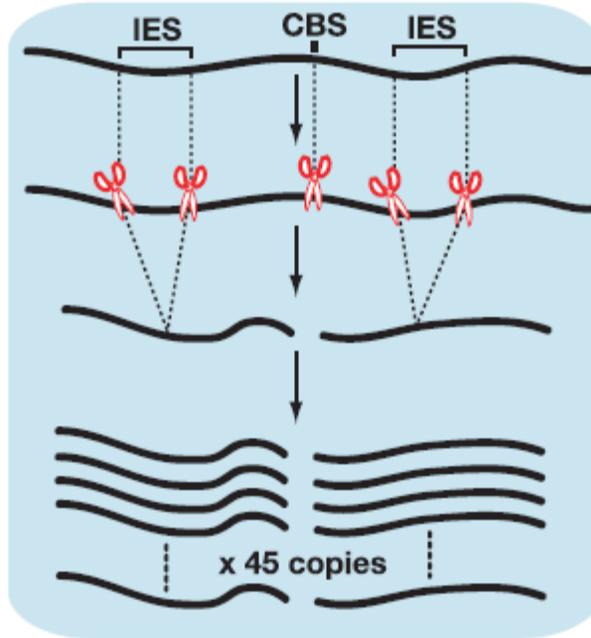


Figure 4. Programmed DNA rearrangements in *Tetrahymena*

Illustration of the result of both breakage eliminated sites (BESs) and internal eliminated sequences (IESs) on *Tetrahymena* macronuclear DNA (adapted from Mochizuki, 2012).

### 1.5.1 Breakage eliminated sites (BESs)

Recognition by unknown endonucleases of highly conserved 15 base pair chromosome breakage sequences (CBSs) results in site-specific chromosome fragmentation (Fan and Yao, 2000). Small regions of DNA flanking the breakage site known as breakage eliminated sequences (BESs) are deleted and replaced with telomere repeats to effectively produce up to 300 macronuclear chromosomes from the original 5 of the micronucleus (Yao *et al.*, 1990).

### 1.5.2 Internal eliminated sequences (IESs)

A second type of rearrangement involves the RNA interference (RNAi) pathway-mediated removal of an estimated 6,000 DNA segments known as internal eliminated sequences

(IESs) ranging in size from 0.5kb to over 20kb (Yao *et al.*, 1984; Mochizuki and Gorovsky, 2004). The non-coding repeat nature of IESs makes their sequences similar to the DNA contained within heterochromatin and transposons, and interestingly, IES removal shares resemblance to RNAi-mediated heterochromatin formation in other eukaryotes (Volpe *et al.*, 2002; Fillingham *et al.*, 2004).

While most eukaryotes retain silenced DNA in dense heterochromatin regions, *Tetrahymena* utilizes an RNA interference (RNAi)-mediated process to remove these sequences completely from the genome in a case of "extreme silencing" (Mochizuki *et al.*, 2002). The mechanisms involved in this large-scale genomic restructuring make *Tetrahymena* a useful model to study large-scale genomic rearrangements and RNA interference (Mochizuki and Gorovski, 2004; Yao and Chao, 2005; Mochizuki, 2010).

Unlike the conserved breakage sites of BESs, IESs show no consistent recognition sequence, although *cis*-acting elements required for efficient programmed elimination have been identified within some IESs such as the mse2.9 IES (Fillingham and Pearlman, 2004) as well as the M and R deletion elements (Godiska and Yao, 1993; Chalker *et al.*, 1999). Studies in *T. thermophila* in which DNA containing an IES was inserted into the parental macronuclear genome inhibited its elimination from the developing macronuclear anlage (Chalker and Yao, 1996). In similar studies in *Paramecium*, when an MDS was deleted in the parental macronucleus, it was also eliminated from the new macronucleus (Epstein and Forney, 1984). The scan RNA (scnRNA) model was proposed to account for the lack of specific IES recognition sequences and to accommodate the epigenetic influence between the old and new MAC, as the micronucleus is the sole progenitor of genetic material (Mochizuki and Gorovsky, 2004).

### 1.5.2.1 Scan RNA model

In this model, the parental macronuclear genome is compared against that of the micronucleus in an effort to identify and remove inequalities from the developing macronucleus. The micronuclear chromosomes are transcribed bidirectionally during early conjugation to form double stranded RNAs (dsRNAs). These dsRNAs are digested at specific but uncharacteristic sites by the Dicer protein Dcl1 to create small RNAs known as scan RNAs (scnRNAs) (Mochizuki and Gorovsky, 2005), which complex with the protein Twi1; a member of the AGO1/Piwi-related protein family commonly involved in RNAi related gene silencing (Mochizuki and Gorovsky, 2004; Cenik and Zamore, 2011). The protein Giw1 transfers the scnRNA-Twi1 pair to the macronucleus (Noto, *et al.*, 2010) whereupon the methyltransferase Hen1 stabilizes the scnRNA (Kurth and Mochizuki, 2009). The scnRNA-Twi1 pair aligns with its complementary macronuclear sequences with the aid of RNA helicase Ema1 (Aronica *et al.*, 2008); an interaction that triggers the degradation of bound scnRNAs. The parental MAC is free of IESs, so only the scnRNAs encoding IES sequences remain. These segments represent the discrepancy between the parental IES-free macronucleus and the developing macronucleus containing IESs, and are transported to the developing macronucleus where they effectively bind and target every IES for elimination.

This interaction induces methyltransferase Ezi1-mediated methylation of H3K9 and H3K27 (Liu *et al.*, 2004b, 2007), which subsequently accumulate chromodomain proteins Pdd1 and Pdd3 in modifications characteristics of heterochromatin (Taverna *et al.*, 2002). The IESs are ultimately excised by the transposase-like protein Tpb2 (Cheng *et al.*, 2010), and flanking macronucleus destined sequences (MDSs) are ligated to maintain the number of chromosomes (Yao *et al.*, 1984).

## **1.6 *Tetrahymena thermophila*: challenges and considerations**

Despite being a strong tool for many applications, there are inevitable challenges that must be taken into account when working with *Tetrahymena*. Its genome composition and unique codon usage impose complications in sequencing and genetic manipulation, and will be discussed forthwith.

### **1.6.1 Highly polyploid macronucleus**

During macronuclear development, a process consisting of multiple rounds of mitosis without subsequent nuclear division known as endocycling replicates the genome to around 45 copies (Eisen, *et al.*, 2006). This highly polyploid transcriptional macronucleus complicates the generation of homozygous recombinants following transformation. Unlike haploid organisms which are homozygous after a single recombination event, *Tetrahymena* require phenotypic assortment to direct homozygosity. The incorporation of the paromomycin-resistance *neo* gene within transformation cassettes enables accelerated phenotypic assortment through manipulation of paromomycin concentration in the *T. thermophila* growth media.

### **1.6.2 AT-rich genome**

In line with the complications derived from a highly repetitive genome, the macronuclear genome of *T. thermophila* is enriched with nitrogenous bases adenine (A) and thymine (T). The average AT content of the macronucleus is 78%, with exonic DNA regions averaging around 72% and non-coding intronic and intergenic sequences peaking upwards of 83% (Eisen *et al.*, 2006). To put this into perspective, humans have an average AT content of approximately 59% (International Human Genome Sequencing Consortium, 2001). It is not the abundance of adenine

and thymine specifically that poses a challenge, but with such a high proportion of the genome consisting of only two nucleotides there is less genetic variance. Simple tasks such as polymerase chain reaction (PCR) face interference when primers binding within introns or intergenic sequences are subject to low sequence complexity, and results in non-specific primer binding and variable products.

### **1.6.3 Codon usage in *Tetrahymena***

Another peculiar feature of the *Tetrahymena* genome is its usage of a non-universal genetic code. In most eukaryotes, three universal stop codons signal the ribosome to terminate translation: UGA, UAG, and UAA. Instead, *Tetrahymena* UGA and UAG encode the amino acid glutamine (Q), while UAA is the only stop codon (Horowitz and Gorovsky, 1985). Although it may be the only stop codon in *Tetrahymena*, genetic evidence in *T. thermophila* uncovered a predicted UGA-decoding tRNA that is specific for the amino acid selenocysteine, which would make it the first organism in which all 64 triplet possibilities had the potential to code for an amino acid (Eisen *et al.*, 2006). The functional consequence of the alternate codon usage in *Tetrahymena* is that if an exogenous gene terminating in TGA or TAG was transformed into *Tetrahymena*, it would be translated beyond its usual termination site and likely result in a misfolded protein product. Likewise, a *Tetrahymena* gene encoding TGA or TAG to specify glutamine would terminate prematurely if transformed into an organism employing canonical codon usage, resulting in a truncated and most likely misfolded protein as well. This can be bypassed when planning gene transformations between *Tetrahymena* and other common laboratory organisms such as bacteria or yeast with the use of synthetic genes that support the universal code, although this is a costly avenue for long genes.

## **1.7 Studying CTD-interacting proteins in *Tetrahymena thermophila***

Improved annotation and advancements in tools to explore *Tetrahymena* are expanding their application within the laboratory to promote new research in exciting ways. The ability to transform the macronucleus with gene tagging or knockout constructs allows researchers to tag a protein with an epitope tag or knock it out completely (Gaertig *et al.*, 1994).

Despite being unicellular, the unique biology of *Tetrahymena* can benefit studies in human related biology in avenues for which the more traditional eukaryotic model systems such as the budding yeast *Saccharomyces cerevisiae* are unable to (Eisen *et al.*, 2006; Collins, 2012). For instance, unique differences between the histone acetylation and chromatin states of the MIC and MAC have enabled exploration into the differences in histone composition and modification between heterochromatin and euchromatin (Vavra *et al.*, 1982). *Tetrahymena* may also be a powerful tool to investigate the consequence of life without a canonical RNAPII CTD, as such a system may offer insight into unique CTD-independent roles of proteins that are not readily tractable in higher eukaryotes possessing canonical CTDs.

## **1.8 Mediator complex**

Transcription is an intricate series of events that is precisely modulated by a plethora of transcriptional regulators enabling transcription of certain genes to be widely up- or down-regulated depending on the individual needs of the cell. Without this fluid regulation, genes would simply be constitutively "on" or "off", leading to over or under production of mRNA. Many transcriptional regulators are species specific as they have evolved concomitantly with the species to meet its unique needs. A few are evolutionarily conserved, however, and constitute a core set of general transcription factors (GTFs) that are necessary and sufficient along with

RNAPII for transcription. The RNAPII class of general transcription factors is comprised of the five general initiation factors TFIIB, TFIID, TFIIE, TFIIIF, TFIIH that are required for transcriptional initiation and promoter release (Roeder, 1996), and Mediator, which plays major roles in post-initiation regulation and are required for diverse developmental pathways (Conaway *et al.*, 2005; Hentges, 2011).

Mediator is an RNAPII CTD-interacting multimeric protein complex consisting of 25 subunits in the budding yeast *Saccharomyces cerevisiae* in which it was discovered and purified (Kim *et al.*, 1994). It was first identified as a necessary constituent for reconstituting RNAPII transcription in a system consisting only of purified RNAPII and general initiation factors, although many of the subunits had previously been identified in yeast mutation screens affecting transcription, and is now considered a classical eukaryotic transcriptional coactivator (Conaway, 2011).

### **1.8.1 Structure and function of Mediator**

Mediator is a fundamental component of the RNAPII transcription machinery. As its name suggests, it mediates communication between gene-specific DNA transactivators, the general transcription factors (GTFs) TFIID and TFIIH, and the CTD of RNAPII (Hengartner *et al.*, 1995; Johnson *et al.*, 2002; Kuras *et al.*, 2003; Esnault *et al.*, 2008; Borggreffe and Yue, 2011). As a CTD-interacting coactivator, Mediator does not directly bind DNA, and instead relays information from enhancer elements to the RNAPII general transcription machinery during transcription through coordinating phosphorylation of the CTD (Myers and Kornberg, 2000). In addition, Mediator also functions in post-initiation regulation including the recruitment of elongation factors and mRNA maturation (Conaway, 2011; Hentges, 2011). Although

facilitating activator dependent transcription is the main duty of Mediator, it can also stimulate RNAPII basal transcription in a system deprived of all but RNAPII and the general initiation factors (Kim *et al.*, 1994).

The overall architecture of the Mediator complex is modular and consists of the head, middle, tail, and kinase or Cdk8 modules (Figure 5). The head module is composed of Med6, Med8, Med11, Med17, Med18, Med19, Med20, and Med22. The function of the head is to bind RNAPII and certain general transcription factors. Since RNAPII is highly conserved between species, proteins of the of Mediator head module display the highest amount of homology between yeast, murine, and human Mediator complexes (Dotson *et al.*, 2000). The middle module is made up of Med1, Med4, Med7, Med9, Med10, Med21, and Med31. The tail module consists of Med2, Med3, Med5, Med15, and Med16, and is the module responsible for binding sequence-specific DNA binding transactivators (Conaway, 2011), so it follows that the proteins composing the tail module of Mediator are divergent (Dotson *et al.*, 2000).

Cyclin-dependent kinase 8 (CDK8), cyclin C, Med12, and Med13 form the kinase module (Borggreffe *et al.*, 2002). Originally, this module was believed only to act in an inhibitory role as its binding to Mediator prevented the recruitment of RNAPII and subsequent transcriptional activation (Hengartner *et al.*, 1995). However, more recent studies have revealed that the kinase module has a stimulatory role in transcription activation *in vitro* and *in vivo* through phosphorylation of the CTD of Rpb1 to promote dissociation of the pre-initiation complex (PIC) and the transition of RNAPII into active transcription (Liu *et al.*, 2004a).

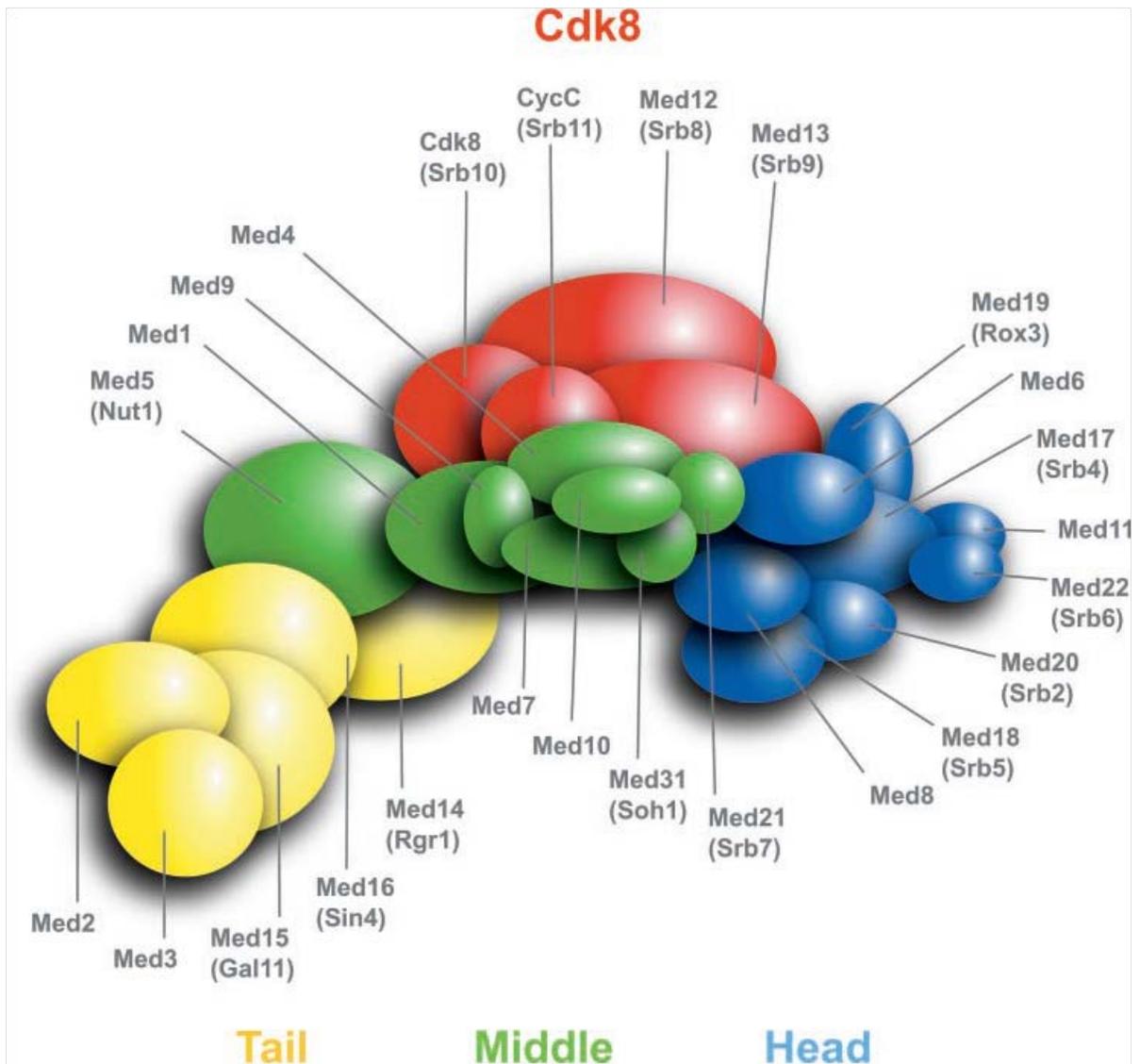


Figure 5. Interaction map of yeast Mediator complex

The Mediator complex is modular and is composed of the Head, Middle, Tail, and CDK8 or kinase module (adapted from Guglielmi *et al.*, 2004).

### 1.8.2 Mediator conservation

Given Mediator's fundamental roles in transcriptional regulation and its close association with RNAPII, it is conserved across animals, fungi, and land plants, and maintains an ancient evolutionary modular organization (Kim, 1994; Fondell *et al.*, 1996; Bourbon, 2008; Mathur,

2011). Orthologs of 22 of the 25 yeast Mediator subunits and 13 *Tetrahymena* genes have been identified among the 33 mammalian subunits (Conaway, 2005; Bourbon, 2008).

Since Mediator is implicated in the regulation of a variety of diverse pathways directing cell growth, development, and differentiation, it is not surprising that ineffectual regulation of its subunits in humans are linked with a variety of human diseases, as well as with specific developmental processes across a variety of organisms (Hentges, 2011; Spaeth *et al.*, 2011). Some subunits are associated with more than one cancer type, such as the overexpression of Med28 (referred to as EG1) linked to cancers of the breast, prostate, and colon (Zhang *et al.*, 2004). In other cases, the same subunit is linked to different types of cancer depending on its molecular disposition. Overexpression of Med1 is associated with breast and prostate cancers, while its underexpression is linked with lung melanoma (Spaeth *et al.*, 2011). Other Mediator-linked human maladies include cardiovascular and neurodevelopment disorders, and even foray into behavioral disorders such as Schizophrenia and psychosis (Spaeth *et al.*, 2011). Interestingly, mutations of genes encoding different Mediator subunits within the same organism generate distinct phenotypes, as do mutations of genes encoding the same subunit within different organisms (Hentges, 2011).

### **1.8.3 Identification of *Tetrahymena* Mediator**

To compare a yeast protein query with a *Tetrahymena* protein database, NCBI's BLASTP was used. Mediator has not been experimentally defined in any protist, but bioinformatic sequence analysis of *S. cerevisiae* Mediator subunits using BLASTP identified Med31 as the sole identifiable Mediator subunit conserved in *T. thermophila* with a remarkable 33% identity over 64% of yeast Med31 ( $E= 1e-11$ ) (Figure 6). In comparison, yeast and human Med31 share 44%

identity with 58% coverage ( $E= 2e-19$ ). More interesting perhaps in lieu of the high conservation is that Med31 is non-essential in either yeast (Fan and Klein, 1994) or *Tetrahymena* (Garg and Fillingham, unpublished); a very unusual phenotype for a gene so evolutionarily conserved across diverse organisms. The high sequence conservation of *MED31* combined with the apparent absence in the genome of other known Mediator subunits, the lack of a canonical CTD of RNAPII, and the non-essential nature of *MED31* in *Tetrahymena* suggests a role independent of classic Mediator, and prompts further characterization of the Mediator complex in *Tetrahymena*.

### **1.9 Integrator complex**

The Integrator complex was initially identified and characterized in human cells based on its co-purification with Dss1; the gene product of a candidate tumor suppressor gene, deleted in split hand/split foot 1, implicated in the human limb malformation split hand/split foot. The gene product of *DSS1* in human cells co-purifies with three major proteins: a Rad51-binding tumor suppressor known as BRCA2 implicated in breast cancer that plays a role in recombination and double-strand-break repair (Marston *et al.*, 1999), the 19S subcomplex of the 26S proteasome involved in the regulation of protein concentrations by degrading unnecessary or damaged proteins (Krogan *et al.*, 2004; Sone *et al.*, 2004), and Integrator; an Rpb1 CTD-interacting protein complex required for the 3' end processing of RNAPII-transcribed uridine-rich (U) snRNAs U1 and U2 destined to become components of the spliceosome (Uguen and Murphy, 2003; Baillat *et al.*, 2005). Integrator therefore links RNAPII transcription with snRNA processing.

### 1.9.1 Structure and function of Integrator

The Integrator complex consists of 12 confirmed subunits in humans (Baillat *et al.*, 2005), but further research in *D. melanogaster* has identified Asunder and CG4785 as additional Integrator subunits with orthologs in humans (Chen *et al.*, 2012).

Transcription of snRNAs continues for about 250bp beyond the coding region to form a 3'-extended snRNA precursor (Cuello *et al.*, 1999). A cleavage specificity sequence known as the 3' box located 9-19bp downstream of the snRNA gene's coding region is necessary for proper 3' end processing (Hernandez, 1985). Integrator is recruited to these non-polyadenylated, 3'-extended pre-snRNA molecules to initiate transcript cleavage just upstream of the 3' box (Baillat *et al.*, 2005) through a unique Ser<sup>2</sup>/Ser<sup>7</sup> double phosphorylation mark on the CTD of Rpb1, which to date remains the only confirmed purpose of Ser<sup>7</sup>P (Egloff *et al.*, 2010). The mature snRNAs bind a number of proteins to form small nuclear ribonucleoproteins (snRNPs), which become the catalytic core of the spliceosome; a multi-megadalton protein complex involved in the removal of introns and alternative splicing during pre-mRNA processing (Will and Lührmann, 2011). Unrelated research into the *Drosophila* Integrator complex identified a requirement for CDK8/cyclin C-independent of Mediator in the 3'-end processing of snRNAs (Chen *et al.*, 2012).

The mechanism behind the 3'-end processing by Integrator is still largely ambiguous, but parallels between Integrator and the cleavage and polyadenylation specificity factor (CPSF) provides some insight. Metazoan CPSF is a protein complex involved in cleaving the 3' end of transcribed pre-mRNA prior to the addition of the characteristic polyadenine tail. CPSF-73 is the specific 3' endonuclease and forms a heterodimeric core complex with CPSF-100 (Mandel *et al.*, 2006; Yang *et al.*, 2009). Homology between most Integrator subunits and other characterized

eukaryotic proteins is minimal, but Integrator subunits Int9 and Int11 bear homology to CPSF-100 and CPSF-73, respectively (Baillat *et al.*, 2005). Interestingly, Int9 and Int11 also undergo a heterodimeric interaction required for snRNA 3' end formation analogous to CPSF-100 and CPSF-73 dimerization (Albrecht and Wagner, 2012). The depletion of Integrator's catalytic subunit Int11 resulted in the accumulation of unprocessed U1 and U2 snRNA transcripts, exemplifying the necessity for this subunit in proficient snRNA 3'-end formation (Baillat *et al.*, 2005). Despite the apparent crossover of function between Integrator and the mRNA 3'-end processing apparatus, it is important to note that the two complexes do not share components and are maintained as separate systems (Ezzeddine *et al.*, 2011).

Further, subunit IntS6 had previously been identified by positional cloning to be the candidate tumour suppressor "deleted in cancer 1" (DICE1) in non-small cell lung carcinoma (Wieland *et al.*, 1999). No other subunits had previously been identified through genetic screens, possibly due to the deleterious nature of these types of studies and the vital requirement of almost all of the Integrator subunits, as observed in *Drosophila* (Baillat *et al.*, 2005; Ezzeddine *et al.*, 2011).

### **1.9.2 Identification of *Tetrahymena* Integrator**

*Tetrahymena* encodes a homolog to human *DSSI*, which prompted an affinity purification that not only identified the 19S regulatory particle of the proteasome and a potential *Tetrahymena* BRCA2 ortholog, but novel peptides indicative of a potential *Tetrahymena* Integrator complex as well (Xiong *et al.*, 2011; Fillingham, unpublished; Table 3). Integrator was believed to be metazoan-specific as no significant homologs could be detected through BLAST search of the GenBank database (Baillat *et al.*, 2005; NCBI). Since *Tetrahymena* encodes

snRNA genes with homology to U1 and U2 snRNA, further investigation is required to characterize this potential snRNA 3'-end processing complex in *Tetrahymena*.

### **1.10 Tandem mass spectrometry and SAINT**

A powerful tool in the realm of proteomics and identification of proteins and protein-protein interactions is tandem mass spectrometry (MS/MS), especially when coupled with affinity purification (AP-MS). The mass spectrometer works by using the difference in mass-to-charge ratio ( $m/z$ ) of ionized protein fragments to separate peptides and quantitate them based on their relative abundance to determine the molecular mass and structural information of the protein. The organization of mass spectrometry is very modular and is essentially composed of three parts: an ionizer, a mass analyzer, and a detector. While different types of these components exist, the setup in the present study utilizes reversed-phase high-performance liquid chromatography (RP-HPLC) coupled with electrospray ionization (ESI) and a linear trap quadrupole (LTQ); abbreviated RP-HPLC-ESI-LTQ-MS/MS.

Reversed-phase high-performance liquid chromatography is a means to separate molecules even before mass spectrometry based on hydrophobicity through differential affinity for the mobile and stationary phases. The sample is loaded into a capillary packed with silica modified with an octadecyl  $C_{18}$  carbon chain to make the silica less polar, and run under high pressure with a mobile phase consisting of 2% acetonitrile, a moderately polar solvent. The more hydrophobic compounds in the sample will bind the silica more often, thus retarding their migration through the capillary. The polarity of the mobile phase is then decreased by adding 98% acetonitrile. This reduces the affinity of the more hydrophobic compounds for modified

silica in favour of the solvent, eluting the remaining proteins. Eluted solvent is directed into the electrospray ionizer.

Awarded a Nobel Prize in 2002 for its discovery (Grayson, 2011), electrospray ionization is a type of ionization source which relies on electricity for ion dispersion. Triggered by expulsion through a charged capillary, the protein sample is converted into an electrically charged mist of mostly positive ions. As the sample solvent evaporates, the charged droplets diminish in size, bringing ions closer together and increasing the surface charge density. Once a critical maximum is reached, the ions can no longer resist their repulsive forces toward one another and eject themselves into the gaseous phase and into the mass analyzer. Since proteins can be relatively large, polar, and thermally unstable, some ionization methods using heat to vaporize the sample would lead to sample fragmentation. ESI is considered a "soft ionization" technique as it does not break any chemical bonds, and for this reason is particularly useful for macromolecules such as proteins (Ho *et al.*, 2003).

Lastly, the linear trap quadrupole (LTQ) is a type of quadrupole ion trap mass analyzer used for trapping and ejecting ions with a particular mass-to-charge ratio on a two-dimensional field. The LTQ grants the ability to select particular ions for another round of mass spectrometry analysis, known as tandem mass spectrometry, without having to run another sample. An associated detector measures the characteristics of the ejected ions. The first MS analysis selects parent ions of certain  $m/z$  ratios for collision-induced dissociation (CID), which fragments the parent ion via collision with neutral molecules. The second MS analysis further evaluates the  $m/z$  ratio of these constituent fragment ions as they are ejected past the detector (March, 1997). The first time each fragment from a single peptide is detected by MS/MS it is counted as a unique

peptide, whereas subsequent identification of any of these fragments counts toward the total peptide number commonly referred to as spectral count (Kean *et al.*, 2012).

The leading computational challenge in mass spectrometry data analysis is filtering both false positives and non-specific protein binding from AP-MS datasets. In situations where protein concentrations are too low to effectively detect using methods such as silver staining, rigorous data analysis becomes more important due to the high sensitivity of mass spectrometry. In attempt to circumvent this, a recently developed computational tool known as Significance Analysis of INTeractome (SAINT) uses label-free quantification and assigns probabilistic confidence scores (0 to 1) to rate the probability that particular bait-prey interactions represent true interactions (Choi *et al.*, 2011). Rather than simply comparing spectral counts between purified and untagged wildtype samples, SAINT uses a statistical algorithm incorporating interaction probability distributions specific to each bait-prey pair from all previously analyzed interactions, and models this data across negative control samples to produce a confidence value (Choi *et al.*, 2011). Ultimately, this method becomes more accurate with each successive analysis, and currently is composed of a database of 16 untagged wildtype controls.

### **1.11 Rationale and project summary**

Although previous affinity purifications coupled with mass spectrometry have identified putative Mediator and Integrator subunits in *T. thermophila*, this organism lacks a canonical CTD on the largest subunit of RNA polymerase II that is necessary for the recruitment of these subunits to genes in humans and yeast. It was therefore my aim to explore the functional consequence of life with a highly divergent CTD through proteomic and genetic analysis of both complexes.

Given the high primary sequence similarity of Med31 between *T. thermophila* and yeast, I transformed a synthetic gene encoding *Tetrahymena MED31* into yeast deleted for *MED31* (*med31Δ*) to determine whether the slow-growth phenotype associated with the *MED31* knockout in *S. cerevisiae* could be rescued, and whether a common function for Med31 existed between the divergent species. In parallel, I explored the prospect of capitulating an interaction between yeast Med8 and *Tetrahymena* Med31 for the purpose of establishing a conserved molecular interaction.

Neither Mediator nor Integrator is expected to exist in *Tetrahymena*, especially given the absence of canonical heptapeptide repeats in its CTD of Rpb1. Therefore, in the second stage of my research, I selected a number of putative *Tetrahymena* Mediator and Integrator subunits for which I engineered epitope tagging cassettes for tandem affinity purification coupled with mass-spectrometry in order to characterize their respective complexes. Lastly, I attempted to knock out putative *Tetrahymena* Integrator subunit *INTS6* through exact gene replacement with a *neo* resistance gene to ascertain whether its gene product is essential or not for cell viability and to enable further investigation into its role in small RNA processing.

## Chapter 2: Materials and Methods

### 2.1 Equipment

All small-scale (1.5ml Eppendorf tube) centrifugations at room temperature were performed using an Eppendorf 5424 centrifuge, while small scale centrifugation at 4°C was carried out in a Sorvall Legend Micro 21R refrigerated microcentrifuge (Thermo Scientific). Centrifugation of 15ml and 50ml Falcon tubes at room temperature was performed with a Centra CL32 (IEC), and refrigerated spins using a Sorvall Legend RT. Large-scale centrifugation (500ml) of *Tetrahymena* cells for affinity purifications was performed with an Avanti J-30I (Beckman Coulter). Polymerase chain reactions (PCR) were performed using a GeneAmp PCR System 9700 (Applied Biosystems) or a GeneAmp PCR System 9600 (PerkinElmer).

### 2.2 Sequence alignments

Sequences for all *S. cerevisiae* proteins were obtained from the Saccharomyces Genome Database (<http://www.yeastgenome.org/>). Sequences for all *T. thermophila* proteins were obtained from the *Tetrahymena* Genome Database ([www.ciliate.org](http://www.ciliate.org/)), see Appendix C.4 for accession numbers. FASTA sequences were aligned using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). The "ALN"-formatted alignment was then run through BoxShade ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)) version 3.21 and presented in RTF-new format to visually enhance alignment similarities.

### 2.3 Growth conditions

*E. coli* transformed with the 12myc-pRb415, pBKS-FZZ, or p4T2-1 vectors were grown overnight on YT plates supplemented with ampicillin (50µg/ml) at 37°C and stored at 4°C. For plasmid preparation, 3ml cultures were grown overnight by shaking at 250rpm at 37°C in YT media supplemented with the appropriate antibiotic. Glycerol stocks were prepared for long-term storage by transferring 0.8ml of overnight liquid culture to 0.8ml of sterile 50% glycerol in a 1.8ml CryoPure cryovial (Sarstedt) and stored at -80°C.

*S. cerevisiae* transformed with the 12myc-pRb415 vector were grown on minimal media plates lacking leucine (YNB -leu) at 30°C and stored at 4°C. For genomic isolation, 3ml cultures were grown overnight by shaking at 225rpm at 30°C in the appropriate media. Glycerol stocks were prepared for long-term storage as above.

*T. thermophila* in liquid culture were grown in flasks containing no more than 1/10 of its total volume of culture. Cultures were grown in the vegetative state overnight in sequestrin proteose peptone (SPP) media at 30°C with shaking at 90rpm. For selection and phenotypic assortment, transformed *Tetrahymena* strains were grown in 96-well microtitre plates (Sarstedt) in SPP supplemented with the appropriate concentration of the antibiotic paromomycin. Starvation was carried out by resuspending log phase cells in 10mM Tris-HCl, pH 7.4 overnight at 30°C without shaking. For freezing cell lines for long-term storage, 100ml of culture were grown overnight in SPP supplemented with penicillin-streptomycin-fungizone (PSF) and then centrifuged at 3,000rpm for 3 minutes at room temperature. Cells were resuspended in 100ml of 10mM Tris pH 7.4 and starved for 2 days at 30°C without shaking in a 1L Erlenmeyer flask. Cells were centrifuged again at 3,000rpm for 3 minutes and aspirated to 250µl. To this, 10% DMSO (Sigma) in 10mM Tris pH 7.4 was immediately added (final DMSO concentration=8%)

and cells were re-suspended. Into individual 1.8ml CryoPure cryovials (Sarstedt), 0.5ml of cells in DMSO were aliquoted and frozen in liquid nitrogen.

#### **2.4 Media, buffers, solutions**

Recipes of all media, buffers, and solutions used are listed in Appendix C.5.

#### **2.5 Manual *T. thermophila* genomic DNA extraction**

Genomic DNA from *T. thermophila* was isolated from wildtype B2086 or CU428 strains as described by Gaertig *et al.* (1994). *Tetrahymena* were grown overnight at 30°C in 3ml SPP+PSF (16 hours). In a 1.5ml Eppendorf tube, 1ml of cells were harvested by centrifugation at 3,000rpm for 2 minutes at room temperature and the supernatant was discarded. The pellet was re-suspended in 500µl *Tetrahymena* lysis solution and mixed until homogenous.

Phenol:chloroform (1:1) extraction was performed twice as follows to remove proteins and lipids: First, 300µl phenol and 300µl chloroform was added to 600µl cell suspension (100µl cell lysate + 500µl lysis solution) and mixed until the solution was homogeneously white and opaque. The mixture was centrifuged at 13,000rpm for 1 minute at room temperature, and the top aqueous layer (~600µl) was transferred to a new Eppendorf tube. Next, a chloroform extraction was performed as follows: 500µl chloroform was added to 500µl of sample (1:1) and mixed until the mixture was homogeneously white and opaque. The sample was centrifuged at 13,000rpm for 1 minute at room temperature and the top aqueous layer was again transferred to a new Eppendorf tube. To 600µl of sample, 200µl of 5M NaCl was added, followed by 800µl of isopropanol to precipitate the DNA. This was then centrifuged at 13,000rpm for 2 minutes at room temperature and the supernatant was removed. The DNA pellet was washed twice: 200µl

of 70% ethanol was added followed by centrifugation at 13,000rpm for 2 minutes and removal of supernatant. The pellet was then desiccated in a vacuum desiccator for 20-30 minutes and then re-suspended in 100µl ddH<sub>2</sub>O before adding 1µl of RNase (10mg/ml) and incubating at 37°C for 1 hour. The sample was left overnight at 4°C to allow further dissolving and stored at -20°C.

## **2.6 Manual *E. coli* plasmid DNA miniprep (alkaline lysis miniprep)**

Plasmid DNA was isolated from *E. coli* using a modified alkaline lysis miniprep method based on that of Birnboim and Doly (1979). *E. coli* were grown overnight at 37°C in 2 ml of YT with appropriate selective drug. In a 1.5ml Eppendorf tube, 1.5ml of culture was harvested by centrifugation at 13,000rpm for 1 minute at room temperature (~21°C) and the supernatant was discarded. The pellet was re-suspended in 100µl of miniprep solution 1 and vortexed until homogenous. Following a 5 minute incubation at room temperature, 200µl of miniprep solution 2 was added and contents were mixed gently to lyse the cells. After another 5 minute incubation on ice, 200µl of miniprep solution 3 was added, and the lysate was placed back in ice for 5 minutes. The sample was then centrifuged at 13,000rpm for 10 minutes to remove cellular debris, and the supernatant was transferred to a new Eppendorf tube. A secondary spin at 13,000rpm for 2 minutes was performed to further remove remaining debris, and the supernatant was transferred again to a new tube. To precipitate DNA, 1ml of 95% ethanol was mixed with the lysate by inversion and allowed to stand at room temperature for 10 minutes. DNA was first pelleted by centrifugation at 13,000rpm for 2 minutes at room temperature, followed by aspiration of the supernatant. The DNA pellet was washed with 200µl 70% ethanol, spun at 13,000rpm for 5 minutes at room temperature, and the supernatant was aspirated away. This wash procedure was repeated once leaving just a DNA pellet which was

placed in a vacuum desiccator for 20-30 minutes to remove residual ethanol. Finally, to solubilize the dried DNA pellet and free it of RNA, it was re-suspended in 50 $\mu$ l of ddH<sub>2</sub>O, treated with 1ml of RNase (10mg/ml), and incubated at 37°C for 1 hour.

## 2.7 Polymerase chain reaction (PCR)

PCR reagents were added to 0.2ml thin-walled PCR tubes as follows (20 $\mu$ l):

1 $\mu$ l genomic DNA, 1 $\mu$ l Forward Primer (30pmol/ $\mu$ l), 1 $\mu$ l Reverse Primer (30pmol/ $\mu$ l), 7 $\mu$ l ddH<sub>2</sub>O, 10 $\mu$ l 2x PrimeSTAR Max DNA Polymerase (TaKaRa).

Samples were run in the following thermal cycler program (30 cycles):

Table 1. PCR conditions for 2x PrimeSTAR Max DNA Polymerase (TaKaRa)

Stage	Temperature	Time
Denaturation	98°C	10 seconds
Annealing	55°C	15 seconds
Elongation	72°C	30 seconds (~5 seconds/kb)

## 2.8 DNA restriction digest/linearization

Enzymatic restriction digests of both plasmid and PCR product DNA were performed in accordance with manufacturer specifications for enzymatic conditions (New England BioLabs), using approximately one unit of restriction enzyme per microgram of DNA. Small-scale diagnostic digests were performed in 10 $\mu$ l total volume for one to two hours, while large-scale digests including preparation of linearized DNA for biolistic applications were carried out in 400 $\mu$ l reactions overnight.

## **2.9 Enzymatic cleanup and gel extraction**

Enzymatic cleanup following PCR was performed using an EZ-10 Spin Column PCR Products Purification Kit (Bio Basic). Cleanup of restriction digests of PCR products or restriction digests requiring gel extraction was performed with an EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic). All handling of DNA was in accordance with manufacturer specifications.

## **2.10 DNA electrophoresis**

Agarose gels (1% w/v) were made with 1X TBE, stained with ethidium bromide (0.1% v/v of 10mg/ml EtBr) for visualization under UV light, and electrophoresed at 100V in 1x TBE buffer. For DNA size determination, 5 $\mu$  of 1kb DNA Ladder (FroggaBio; Appendix C.6), range 250bp to 1,000bp was used as a standard molecular weight size marker. DNA samples were mixed with 6x DNA loading dye just prior to loading into agarose gel. Visualization was performed under UV light using either a Bio-Rad XRS+ Imager at Ryerson University or an Alpha Innotech FluorChem system in the Core Molecular Facility at York University.

## **2.11 DNA ligation and transformation into competent *E. coli***

Molecular cloning was carried out in *E. coli* grown in YT supplemented with antibiotic corresponding to the plasmid's resistance gene. Vectors 12myc-pRb415, pBKS-FZZ, and p4T2-1 all contain an ampicillin resistance gene (Amp<sup>R</sup>), and were thus grown in YT+amp (50 $\mu$ g/ml). Ligation reactions were performed in 10 $\mu$ l total volume using 1 $\mu$ l of T4 DNA Ligase (NEB). The ligation reaction was incubated at room temperature for 2 hours, after which 25 $\mu$ l high-efficiency competent *E. coli* cells. (NEB 5-alpha, New England BioLabs) were added and transformation

followed the "High Efficiency Transformation Protocol" for "C2987" provided by New England BioLabs. The cells were incubated on ice for 30 minutes and then heat-shocked in a 42°C water bath for 30 seconds. After an additional 5 minute incubation on ice, 950µl room temperature SOC media was added and cells were shaken at ~250rpm for 1 hour at 37°C. Cells were pelleted by centrifugation at 13,000rpm for 30 seconds, and all but 100µl was aspirated. All 100µl was plated on selective media and grown overnight at 37°C. As a negative control, ddH<sub>2</sub>O was substituted for DNA in the transformation reaction to ensure that competent cells were not naturally antibiotic resistant.

## **2.12 Sequencing**

Sequencing of 5' and 3' homology sequences was performed by ACGT Corporation, The Centre for Applied Genomics (MaRS, The Hospital for Sick Children), and Lee Wong (Sequencing Facility, York University). For a list of sequencing primers see Appendix C.7.

### **2.12.1 *E. coli* plasmid DNA isolation for sequencing**

DNA was isolated from 3ml cultures of *E. coli* grown overnight at 37°C using a High-Speed Plasmid Mini Kit (GeneAid) and an EZ-10 Spin Column Plasmid DNA Minipreps Kit (Bio Basic) in accordance with manufacturer specifications.

### **2.13 DNA purifications**

For the purpose of *Tetrahymena* biolistics, DNA was usually purified using an EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic). When product yield was insufficient, DNA purification by either 13% PEG or ethanol precipitation was utilized.

#### **2.13.1 PEG purification**

Following enzymatic digestion, an equal amount (1:1) of 13% polyethylene glycol (PEG) + 1M NaCl was added to the sample and incubated overnight at 4°C. Then it was centrifuged at 13,000rpm for 30 minutes at 4°C and the supernatant was discarded. The remaining DNA was washed with 500µl 70% ethanol and centrifuged at 13,000 rpm for 5 minutes at room temperature (~21°C), again discarding the supernatant. Desiccation of the DNA was carried out in a vacuum desiccator for 20-30 minutes. The DNA pellet was re-suspended in 20µl ddH<sub>2</sub>O and 1µl of the final product was run on a 1% agarose gel to determine DNA concentration.

#### **2.13.2 Ethanol purification**

Following enzymatic digestion, 1/10 volume (40µl in 400µl) of 3M NaOAc pH 5.2 was added to the sample followed by 2 volumes of 95% ethanol (800µl) at room temperature. Incubation was carried out either overnight at -20°C or 30 minutes at -80°C, followed by centrifugation at 13,000rpm for 15 minutes at 4°C. After discarding the supernatant, the DNA pellet was washed with 1ml of ice-cold 70% ethanol, dislodging the pellet, and centrifuged at 13,000 for 15 minutes at 4°C. The supernatant was discarded and the sample desiccated for 20-30 minutes in a vacuum desiccator. The dried DNA pellet was re-suspended in 20µl of ddH<sub>2</sub>O, and 1µl of final product was run on a 1% agarose gel to determine DNA concentration.

## 2.14 Construction of the 3xFLAG-TEV-ZZ (FZZ) tagging cassette

The tagging cassette consisted of a triple FLAG tag fused to a tobacco etch virus cleavage site fused to two repeats of the Z domain of the *Staphylococcus aureus* protein A (3xFLAG-TEV-ZZ, or FZZ), and was designed to tag specific target genes at the C-terminus. The cassette also contained a *neo2* gene, which endowed the transformed *Tetrahymena* with resistance to aminoglycosides such as paromomycin, a protein synthesis inhibitor (Mehta and Champney, 2003). The benefit of using *neo* is that it is not endogenous to *Tetrahymena* so there is no homologous site into which it can recombine.

Directional cloning into the pBluescript vector of both 5' coding "5' homology sequence" and 3' untranslated region (3'UTR) "3' homology sequence" that were homologous to regions flanking the stop codon was required for each target gene. This ensured proper orientation of the linearized tagging cassette relative to the endogenous sequence into the *Tetrahymena* macronuclear genome during homologous recombination. The 5' sequence was constructed by PCR amplification of *Tetrahymena* genomic DNA using an upstream forward primer (UF) complimentary to 30bp of sequence 1kb 5' (upstream) of the stop codon of the target gene, and an upstream reverse primer (UR) complimentary to 30bp immediately upstream of the same stop codon. The 3' sequence was similarly constructed by PCR amplification of *Tetrahymena* genomic DNA using a downstream forward primer (DF) complimentary to 30bp immediately downstream of the target gene's stop codon and a downstream reverse primer (DR) complimentary to 30bp of sequence 1kb 3' (downstream) of the stop codon. The UF primer contained a *KpnI* restriction site while the UR primer contained a *XhoI* restriction site so the final 1kb PCR product had flanking restriction sites and could be digested and cloned into the

complimentary digested FZZ vector. Likewise, the downstream forward (DF) and reverse (DF) primers contained *NotI* and *SacI* restriction sites, respectively (Appendix C.8).

Diagnostic digests were performed after each cloning step to assess proper insertion of the homology sequence by analyzing the size of the digested fragments. Since the 5' and 3' homology sequences were around 1kb long, a digested fragment of this size indicated a potential correct insertion. Sequencing of both 5' and 3' homology sequences was carried out to confirm their identity and to verify that the 5' sequence was in frame with the tag. The primer M13R was used for sequencing into the 5' sequence. When insufficient sequencing data was returned to confirm in-frame fusion with the FZZ tag, primer HN111, which binds the tagging region, was utilized for sequencing across the 5' sequence/FZZ junction to ensure in-frame fusion. The 3' sequence was sequenced with the M13F primer, which sequences from the backbone of the plasmid across the junction into the 3' homology sequence. The DNA sequences of all sequencing primers are available in Appendix C.7.

Lastly, linearization of the tagging cassette prior to biolistic transformation was achieved by digesting the vector with *KpnI* and *SacI* restriction enzymes. The digestion products were separated on an agarose gel to ensure two bands were observed; one matching the 3kb vector and the other matching the 4.5kb tagging cassette.

### **2.15 Construction of the knockout cassette**

The knockout cassette involved two-step directional cloning into the p4T2-1 vector. Primers were designed to amplify 1kb regions immediately upstream of the start codon (3' UTR) and immediately downstream of the stop codon (5'UTR) of the target gene using *T. thermophila* genomic DNA as a template. Cloning was verified by diagnostic restriction enzyme digest, and

sequencing was achieved using the primer H4neo reverse (H4neoR), which binds the H41 promoter region of the cassette for sequencing across the junction into the 5' sequence, and primer BTU2R which binds the BTU2 region of the cassette for sequencing across the junction into the 3' sequence. Linearization was performed as above with *KpnI* and *SacII* restriction enzymes.

## **2.16 Biolistic transformation of *T. thermophila***

Macronuclear exact gene replacement by particle bombardment and homologous recombination was achieved through biolistic transformation of *Tetrahymena* as described by Bruns and Cassidy-Hanley (2000) using a biolistic PDS-1000/He Particle Delivery System (Bio-Rad). This system uses helium pressure to accelerate DNA-coated gold particles into *Tetrahymena*.

### **2.16.1 Preparation of *T. thermophila* cells**

Two days before the planned transformation, 50ml of B2086 or CU428 wildtype *Tetrahymena* cells per transformation were grown in sterilized 500ml Erlenmeyer flasks by shaking overnight at 30°C in SPP+PSF. All 50ml were collected by centrifugation for 3 minutes at 5,000rpm at room temperature in 50ml Falcon tubes, washed with 50ml of 10mM Tris pH 7.4, and centrifuged again. After discarding the supernatant, the cell pellet was re-suspended in 10mM Tris pH 7.4 and transferred into a new sterilized 500ml Erlenmeyer flask and starved for 16 hours at 30°C. Cells were collected again by centrifugation, this time at 1,000rpm for 5 minutes at room temperature as cells are under stress following starvation. All but 0.5-1ml of supernatant was aspirated and the cells were re-suspended in the remaining supernatant.

### **2.16.2 Preparation of gold beads**

All gold preparation steps should be performed at 4°C. To prepare 1.0µm gold beads (Bio-Rad) for transformation, bath sonication to separate individual gold particles was followed by aliquoting 25µl of gold beads per transformation into low-retention 1.5ml Eppendorf tubes (Axygen). Under constant vortexing, constituents were added in the following order to precipitate DNA onto the beads: 3-4µl of linearized tagging-cassette DNA (1µg/µl), 25µl of ice-cold 2.5M calcium chloride (CaCl<sub>2</sub>), and 10µl of cold 0.1M spermidine (Sigma). The gold pellet was then washed twice and pulsed down in the centrifuge at 14,000rpm for 5 seconds between each wash. The first wash was with 200µl of ice-cold 70% ethanol, and the second was with 200µl of ice-cold 100% ethanol. Lastly the pellet was re-suspended in 20µl of cold 100% ethanol.

### **2.16.3 Preparation of flying discs**

Flying discs (Bio-Rad) were dipped in 95% ethanol and let air dry in the laminar hood before being inserted into the ring holder. The 20µl of prepared gold beads+DNA were bath sonicated and vortexed, and spotted onto the middle of the disc. The disc and holder were placed into a desiccator to dry the gold.

### **2.16.4 Assembly and operation**

The gene gun parts were washed first with ddH<sub>2</sub>O and then 70% ethanol to sterilize and the helium pressure was set to 1,100psi. The gene gun was assembled from the top down. A 900psi rupture disc (Bio-Rad) was dipped in isopropanol and fitted into the holder which was screwed into place with a torque tool. Next, the flying disc and holder were inverted and placed

on top of a stopping screen (Bio-Rad) which was fitted into the flying disc platform and secured with a screw cap. This assemblage was inserted into the top slot of the gene gun. A piece of Whatman filter paper and 2ml of 10mM Tris pH 7.4 was added to the lid of a Petri plate, whereupon the 0.5-1ml of prepared *Tetrahymena* cells for transformation were added and placed in the second slot from the bottom in the gene gun.

Vacuum pressure within the gun was allowed to reach 25-26psi, and the starved *Tetrahymena* cells were bombarded with the DNA-coated gold beads at around 900psi. The filter paper and remaining buffer were transferred to 500ml Erlenmeyer flasks containing 50ml of pre-warmed (30°C) SPP+PSF and incubated at 30°C for 4 hours with shaking to recover.

#### **2.16.5 Selection**

Following incubation, all 50ml of cells were centrifuged in a 50ml Falcon tube at 1,000rpm for 5 minutes and all but 15ml of supernatant was aspirated. The pellet was re-suspended and 12µl of 100mg/ml of paromomycin was added to a final concentration of 80µg/ml paromomycin. All 15ml were plated in a 96-well microtitre plate, 200µl per well, and incubated at 30°C for 72 hours. Cells were assessed for robust growth before being put on a regimen of increasing paromomycin concentrations.

#### **2.17 Rapid *T. thermophila* DNA extraction/colony PCR**

*Tetrahymena* were grown overnight by shaking at 30°C in 1ml SPP+PSF (16 hours). To 50µl of cells, 500µl Proteinase K buffer was added in a 1.5ml Eppendorf tube and incubated in a 55°C water bath for 1 hour. Samples were then boiled for 10 minutes and incubated on ice for 5 minutes. For PCR purposes, 5µl was used per PCR reaction and run at previous conditions.

## **2.18 Preparation of *T. thermophila* cell extracts (TCA extraction)**

*Tetrahymena* were grown overnight by shaking at 30°C in 3ml of SPP+PSF (16 hours). In 1.5ml Eppendorf tubes, 1ml of culture was harvested by centrifugation at 13,000rpm for 3 minutes at room temperature (~21°C). After discarding the supernatant, cells were washed once with 1ml 10mM Tris pH 7.4 and spun again at 13,000rpm for 3 minutes. The total volume was aspirated down to 100µl, and 10µl of 100% trichloroacetic acid (TCA) was added and mixed by inversion (final TCA concentration = 10%). The sample was placed on ice for 30 minutes and centrifuged at 12,000rpm for 2 minutes at room temperature. After discarding the supernatant the cell pellet was re-suspended in 100µl 2x SDS buffer. At this stage the solution sometimes turned yellow due to high acidity. To counteract this, 1µl at a time of 1N NaOH was added until the original dark blue colour was restored. The sample was boiled for 5 minutes and incubated on ice for 2 minutes. The supernatant was transferred to a fresh 1.5 ml Eppendorf tube, centrifuged at 13,000rpm for 30 seconds, and kept on ice until gel loading.

## **2.19 Western blot analysis**

To observe target proteins in a heterogeneous protein sample, they were electrophoretically separated by size on a sodium dodecyl sulfate (SDS) polyacrylamide gel, transferred to a polyvinylidene fluoride (PVDF) membrane, and probed with antibodies in a 2-step detection process.

### **2.19.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)**

SDS-PAGE was used to denature and separate proteins based on size. The polyacrylamide gel was composed of 1ml of 5% stacking gel was layered on top of 5ml of 10%

running gel using a Mini-PROTEAN Tetra Cell Casting Module (Bio-Rad) using a 1mm comb for well formation. Sample loading consisted of 5 $\mu$ l of PiNK Plus Prestained Protein Ladder (Genedirex; Appendix C.6), range 10-175kDa, and 20 $\mu$ l of prepared protein sample. The gel was electrophoresed in 1x Western buffer in a Mini-PROTEAN Tetra Cell (Bio-Rad) at a constant voltage of 100V.

### **2.19.2 Western transfer**

A PVDF membrane (Bio-Rad) was prepared by washing it in 100% methanol for 5 minutes, a rinse in ddH<sub>2</sub>O for 3 minutes, then washed in 1x Western transfer buffer for 3 minutes. For wet transfer of proteins, a "sandwich" saturated in 1x Western buffer was set up in the following order from cathode (negative) to anode (positive): a sponge, two pieces of Whatman filter paper, the SDS-PAGE gel, one PVDF membrane, two more pieces of Whatman paper, lastly followed by another sponge. Transfer was carried out in a water-cooled gel running apparatus filled with 1x Western transfer buffer at 80V for one hour or 15V overnight. For semi-dry transfer, a Trans-Blot SD semi-dry electrophoretic transfer cell (Bio-Rad) was used. Layers were stacked bottom (anode) to top (cathode) as follows: two pieces of Whatman paper, PVDF membrane, SDS-PAGE gel, followed by two more pieces of Whatman paper, all saturated in Western transfer buffer. No additional buffer was used in this setup, and transfer was run at 10V for one hour.

### **2.19.3 Ponceau stain**

The PVDF membrane was incubated with 0.1% Ponceau stain for 5 minutes and de-stained with 1x PBS to check for proper transfer of proteins. The membrane was de-stained with three 10-minute washes in 1x PBS solution at room temperature.

### **2.19.4 Blocking**

To reduce non-specific binding of subsequent antibody probes, blocking was performed post-transfer by incubating the blot in 5% milk solution (BLOTTO) for one hour at room temperature or overnight at 4°C. Three 10-minute washes were performed in 1x PBS at room temperature to remove unbound milk protein.

### **2.19.5 Probing**

Unless otherwise noted, Western blots were incubated in monoclonal mouse  $\alpha$ -FLAG (Sigma-Aldrich) primary antibody diluted 1:500 in 5% milk solution, or monoclonal mouse  $\alpha$ -actin (GenScript) primary antibody diluted 1:1000 in 5% milk, for one hour at room temperature or overnight at 4°C. Washing was carried out three times in 1x PBST for 10 minutes each at room temperature. Incubation with horseradish peroxidase-conjugated polyclonal goat  $\alpha$ -mouse (Cedarlane) secondary antibody diluted 1:3000 in 1% milk solution for one hour at room temperature or overnight at 4°C. Visualization was achieved by incubating the blot in Luminato Crescendo Western HRP Substrate (Millipore) according to manufacturer specifications, and developed on HyBlot autoradiography film (Denville Scientific).

## **2.20 Tandem affinity purification (TAP) in *T. thermophila***

Tandem affinity purification (TAP) was performed to isolate FZZ-tagged proteins in *Tetrahymena* along with their interacting partners.

### **2.20.1 Growing large cultures of *T. thermophila***

Sterilized 250ml Erlenmeyer flasks containing 25ml SPP+PSF were inoculated with 1ml of overnight *Tetrahymena* culture and grown for another 16 hours (overnight) at 30°C. The 25ml of culture were then added to 750ml of SPP+PSF in sterile 1L glass vessels and aerated by bubbling overnight at 30°C. Harvesting was done in 500ml plastic screw-lid containers by centrifugation at 5,000rpm for 5 minutes at 4°C. Cells were washed twice with 10mM Tris-HCl, pH 7.4, and concentrated into 50ml Falcon tubes upon the second wash. Cells were centrifuged once more at 5,000rpm for 3 minutes, all supernatant was aspirated away, and the resulting 3-5ml cell pellet was immediately frozen at -80°C for storage until ready for TAP.

### **2.20.2 Preparation and clarification of *T. thermophila* whole cell extracts (WCEs)**

Frozen cell pellets were thawed by alternately rotating the tubes by hand in room temperature water for 30 seconds and placing them on ice for 30 seconds, and then stored on ice. At this time, 500µl of protease inhibitors (Sigma) and 200µl of 100mM PMSF made in isopropanol was added to 25ml of 2x lysis buffer. Each cell pellet was resuspended in an equal volume of ice-cold 2x lysis buffer+protease inhibitors and then adjusted to 15ml with 1x lysis buffer+protease inhibitors. To this, 300µl of 10% NP-40 (final [C]=0.2% v/v) and 5µl of benzonase nuclease (Sigma) was added. Tubes were rotated end-to-end for 1 hour at 4°C to allow benzonase to digest all released genomic DNA.

The 15ml of whole cell extracts (WCEs) were then divided into 10 separate 1.5ml Eppendorf tubes and centrifuged at 13,000rpm for 30 minutes at 4°C. The supernatant was transferred to new Eppendorf tubes and the centrifugation was performed again to remove remaining debris. Clarified WCEs were then pooled into 15ml Falcon tubes.

Affinity purifications were recurrently afflicted by what was believed to be insoluble lipids and carbohydrates in the elute. This appeared to interfere with IgG binding or TEV cleavage as samples with the characteristic cloudy supernatant during immunoprecipitation would consistently produce poor final elute signals in Western blots. Prevention was seemingly by chance since the cause was not certain, however, overgrown cell cultures would present this most often. In attempt to circumvent the negative effects, overnight cell cultures were often grown for less time or with less inoculation volume, cell pellets were washed and aspirated more vigorously prior to starvation, and whole cell extracts were subject to an additional low speed centrifugation at 5,000rpm.

Following the last centrifugation, 100µl of WCE was transferred to a fresh Eppendorf tube and immediately stored at -80°C to be used as input material for affinity purification. The remaining ~15ml of WCE was added directly to the 50ml Falcon tubes containing 250µl of IgG-Sepharose slurry (see below) and rotated end-to-end for 4 hours at 4°C.

### **2.20.3 Preparation of IgG-Sepharose**

IgG-Sepharose beads (GE Healthcare) were supplied in 0.05M sodium phosphate and 20% ethanol as preservative, and had to be washed immediately prior to use. Each purification required 250µl of suspended IgG-Sepharose beads, so for four tubes, 1ml was transferred to a 1.5ml Eppendorf tube (cut P1000 tips were used for all handling and transferring of IgG-

Sepharose beads). The beads were washed three times by centrifugation at 2,000rpm for 2 minutes at 4°C, followed by aspiration of the supernatant and resuspension of the beads in 500µl of 1x lysis buffer to create a 1:1 beads: buffer slurry. Four 50ml Falcon tubes were set up on ice, and 250µl of slurry was added to each.

#### **2.20.4 TEV cleavage**

Following 4 hour rotation with IgG-Sepharose, samples were centrifuged at 4,000rpm for 5 minutes at 4°C. The beads were washed once with 20ml of IPP300 and twice with 15ml of 1x TEV buffer. During each wash, tubes were rotated for 5 minutes at 4°C, centrifuged at 3,000rpm for 2 minutes at 4°C, and the supernatant was carefully aspirated. After the final aspiration, the beads were transferred to separate 1.5ml Eppendorf tubes along with 750µl of 1x TEV buffer and washed once more as above in 1ml of 1x TEV buffer. Beads were then resuspended in an equal volume of 1x TEV buffer, and 8µl of TEV protease (2mg/ml; gift from J. Greenblatt lab) was added to cleave the TEV cleavage site of the FZZ epitope tag. This frees the protein from the two A domains and the bound IgG-Sepharose. The tubes were wrapped in parafilm to prevent leakage and rotated end-to-end overnight at 4°C.

Following overnight incubation, samples were centrifuged at 2,000rpm for 2 minutes at 4°C, and the supernatant (TEV elute) was transferred to individual ice-cold 1.5ml Eppendorf tubes. To collect all remaining protein, 600µl of IPP100 was added to the remaining beads. This was then mixed by inversion, centrifuged as above, and the supernatant was added to the previous supernatant in the ice-cold 1.5ml Eppendorf tubes.

### **2.20.5 Preparation of M2-agarose**

For each purification, 30µl of M2-agarose was used. Since this is packaged as a 1:1 mixture with ethanol, 60µl of slurry was required for each purification. To prepare M2-agarose for four samples, 300µl of slurry was centrifuged at 3,000rpm for 2 minutes at 4°C. The supernatant was aspirated away and the pellet was resuspended in 1ml of IPP100 buffer with rotation. This process of washing and aspirating was repeated four times to ensure all ethanol is washed away. The final pellet was resuspended in an equal volume of IPP100 buffer, which was around 150µl, and 60µl of this slurry was aliquoted into four pre-chilled Eppendorf tubes. The TEV elute was added to the slurry and the tubes were rotated end-to-end for three hours at 4°C.

### **2.20.6 Final elution**

Following the 3-hour M2-agarose bead binding, samples were washed twice as above (centrifuged at 3000rpm for 2 minutes at 4°C, aspirated, and resuspended in 1ml of buffer) with IPP100 buffer, and then twice in IPP100 buffer made without NP-40 detergent. The pellet was then washed with 750µl of 2mM CaCl<sub>2</sub>/20mM Tris and centrifuged at 5000rpm for 1 minute at room temperature. After removing the supernatant, the pellet was rotated with 500µl of 0.5M NH<sub>4</sub>OH and rotated at room temperature for 20 and then centrifuged at 5000rpm for 2 minutes at room temperature. The supernatant was transferred into new pre-chilled Eppendorf tubes and frozen at -80°C to be used for Western Blots and mass spectrometry.

### **2.21 Mass spectrometry**

Affinity purification coupled to mass spectrometry (AP-MS) is an effective and broadly applicable method for exploring protein-protein interactions. Reversed-phase high-performance

liquid chromatography electrospray ionization linear trap quadrupole tandem mass spectrometry (HPLC-ESI-LTQ-MS/MS) was carried out by Dr. Lambert in the Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital. Following tandem affinity purification (TAP), purified proteins eluted in 0.5M NH<sub>4</sub>OH were centrifugally evaporated in a vacuum concentrator (SpeedVac). The proteins were digested with trypsin; a serine protease which cleaves peptide bonds C-terminal of lysine (K) and arginine (R). Digested protein fragments were pneumatically "bomb-loaded" into a 75µm capillary column packed with silica modified with an octadecyl C<sub>18</sub> carbon chain. This reversed-phase high-performance liquid chromatography column was placed in-line with a linear trap quadrupole (LTQ) mass spectrometer via an electrospray ionization delivery system. The ionized species were analyzed by tandem mass spectrometry (MS/MS). Data-dependent acquisition limited tandem mass spectrometry to the 5 most abundant ions per peak "centroid" and the resulting data files were analyzed by the statistical evaluation program Mascot version 2.3 against the *T. thermophila* RefSeq protein database (NCBI). The fragment mass tolerance was 0.6Da with an average mass window of +/- 3Da. The ion score cutoff parameter was set to 35 and a protein was considered a hit only if it produced two "bold, red" peptides; red indicating the peptide was the top scoring match, and bold indicating the protein was the highest scoring match the peptide was found in. To rate the probability of identified proteins and identify significant interactions, SAINT analysis by SAINT version 3 was performed. Each peptide identified is novel, and had not been previously identified in mass spectrometry data from 16 wildtype purifications (J.P. Lambert, personal communication, August 14, 2013).

## 2.22 Transformation of *S. cerevisiae*

Gene replacement by homologous recombination was performed as described by Schiestl and Gietz (1989). After growing an inoculation of yeast in 2mls of YPD overnight at 30°C, cells were diluted to an optical density of 0.2OD in 100ml of YPD and grown at 30°C until an OD of approximately 1.0 was achieved. Cells from 50mls of media were harvested in a 50ml Falcon tube by centrifugation at 4,000rpm for 4 minutes at 4°C. Cells were then washed once with 10mls of sterilized ddH<sub>2</sub>O and centrifuged again as above. After removal of the supernatant, the cell pellet was re-suspended in 1ml of 100mM lithium acetate (LiOAc), transferred to a 1.5ml Eppendorf tube, and centrifuged at 13,000 for 10 seconds. Since there was approximately 100µl of cells after removal of the supernatant, 400µl of 100mM LiOAc was added to re-suspend the cells in a 4:1 ratio of LiOAc:cell pellet. Each transformation reaction requires 50µl of this cell slurry, which was pipetted into separate Eppendorf tubes and centrifuged at 13,000rpm for 10 seconds so lithium acetate could be aspirated away.

To the pellets, 240µl of 50% polyethylene glycol (PEG) was gently added to protect the cells from the permeabilizing effect of lithium acetate. Next, 36µl of 1M LiOAc pH 7.2 was layered on top, followed by 50µl of 2mg/ml salmon sperm carrier DNA (10mg/ml, Invitrogen) which was denatured prior to use by boiling for 5 minutes. If the PCR product was being transformed, 50µl of PCR product was added next, otherwise 2µl of plasmid DNA in 48µl of ddH<sub>2</sub>O was added. For a negative control, 50µl of ddH<sub>2</sub>O was added instead of plasmid DNA to ensure the cells could not produce leucine without plasmid transformation. Digested empty vector was also transformed (2µl in 48µl of ddH<sub>2</sub>O) as a control for self-ligation of the digested plasmid and for downstream affinity purification purposes to ensure the vector itself does not have affinity for the target protein. The tubes were vortexed until the cell suspension was

homogenous, and then placed at 30°C for 20 minutes. Following incubation, cells were centrifuged at 13,000rpm for 10 seconds at room temperature (~21°C) and the supernatant is removed. Finally, the cell pellet was re-suspended in 50µl of ddH<sub>2</sub>O. If the cells were being transformed with 12myc-pRb415, this entire volume was spread-plated onto YNB -leu minimal media lacking leucine. If the cells were being transformed with pFA6a-13myc-kanMX6 PCR product, transformed cells were first plated onto non-selective YPD media overnight at 30°C, and then replica plated onto YPD supplemented with geneticin (G418) sulfate and incubated at 30°C for 48 hours.

### **2.23 Engineering of yeast and *Tetrahymena* *MED31* expression vectors**

Both *MED31*<sup>Sc</sup> and synthetic *MED31*<sup>Tt</sup> were cloned into the CEN-based plasmid 12myc-pRb415 (Amp<sup>R</sup>/Leu<sup>+</sup>) with *Bam*HI and *Sall* restriction endonucleases and transformed into NEB 5- α competent *E. coli* to tag Med31 with an N-terminal 12xmyc epitope tag.

*MED31*<sup>Sc</sup> was previously knocked out in the yeast Med8-TAP strain (OB<sup>+</sup>/Med8-TAP), originally obtained from the Open Biosystems yeast TAP-fusion library (Thermo Scientific). The strain used to construct this library was Open Biosystems S288C, designated OB<sup>+</sup> for short. Minipreparations of 12myc-*MED31*<sup>Sc</sup> and 12myc-*MED31*<sup>Tt</sup> were transformed into the yeast strain Med8-TAP/*med31*<sup>Sc</sup>Δ (See Materials and Methods). Empty digested 12myc-pRB415 vector (+P) was transformed into Med8-TAP/*med31*<sup>Sc</sup>Δ and OB<sup>+</sup>/Med8-TAP as negative controls for growth and affinity purification. Transformants were plated on YPD media for recovery and incubated for 48 hours at 30°C and then replica plated onto minimal -leu media and incubated for 1-2 days at 30°C for selection.

## 2.24 Affinity purification in *S. cerevisiae*

Affinity purification in yeast was performed using a modified one-step process based on that of Keogh *et al.* (2006). Purification relies on the affinity of the TAP epitope tag for immunoglobulin G (IgG). The TAP tag is composed of three sections: the calmodulin binding peptide (CBP), a tobacco etch virus (TEV) protease cleavage site, and protein A from *Staphylococcus aureus*, with the TEV site in the middle linking CBP and protein A. During affinity purification, IgG recognizes and binds to protein A, allowing "pull down" from solution of recombinant protein fused to TAP tag along with all associating proteins.

After growing an inoculation of yeast in 10mls of YPD overnight at 30°C, cells were diluted to an optical density of 0.2OD in 200ml of YPD and grown at 30°C until an OD of approximately 1.0 was achieved. Cells are harvested by centrifugation at 4,000rpm for 4 minutes at 4°C in 50ml Falcon tubes. Next, cells were washed once with 10mls of ice cold AP wash buffer and centrifuged again as above before removing the supernatant. At this stage, cell pellets can be frozen in liquid nitrogen or dry ice and ethanol and stored at -80°C.

Each purification requires 25µl of packed IgG-sepharose beads (GE Healthcare), and since they are supplied 1:1 in 20% ethanol, 50µl of slurry was used per purification. Beads were pipetted using a cut P1000 tip into a 1.5ml Eppendorf tube, and then centrifuged at 2,000rpm for 2 minutes at 4°C for removal of the ethanol. The beads were then washed three times by adding 750µl cold AP wash buffer, mixing gently by inversion, centrifuging at 2,000rpm for 2 minutes at 4°C, and discarding supernatant. After the final wash, AP wash buffer was added in a 1:1 ratio to packed bead volume, the slurry was mixed, and 50µl was aliquoted to 1.5ml Eppendorf tubes using a cut P200 tip.

Cells were thawed on ice and re-suspended in 850 $\mu$ l of AP lysis buffer and transferred into 2ml conical tubes containing 1ml of glass beads. Cell lysis was performed via bead beating in a Mini-BeadBeater-16 (BioSpec Products) using a regimen of beating for 30 seconds at room temperature followed by 2 minutes on ice for 7 cycles, as deemed experimentally optimal for *S. cerevisiae* by my own analysis. Holes were poked through the bottom of the plastic conical tubes with a 25Gx5/8 in. PrecisionGlide syringe (BD) and whole cell extracts were filtered into a 15ml Falcon tube by centrifugation at 4,400rpm for 1 minute at 4°C. Pellets were re-suspended and transferred to 1.5ml Eppendorf tubes before being centrifuged at 14,000rpm for 20 minutes at 4°C. The supernatant was transferred to a new Eppendorf tube, and 25 $\mu$ l was stored at -80°C as an input control. The remaining supernatant (~800 $\mu$ l) was transferred to a tube containing the IgG-sepharose beads and AP wash buffer, and the tubes were rotated for 2 hours at 4°C. Following rotation, the beads were, then washed three times by adding 750 $\mu$ l cold AP wash buffer, rotating for 5 minutes at 4°C , centrifuging again at 2,000rpm for 2 minutes at 4°C, then discarding the supernatant. During one of these washes, beads and buffer were transferred to a new Eppendorf tube to separate from beads potentially bound to the tube. After the last wash, all AP buffer was removed and 50 $\mu$ l of 2x SDS buffer was added, while 25 $\mu$ l of 2x SDS was added to input samples. Samples were boiled for 5 minutes whereupon they were ready to be run on SDS-PAGE gels or stored at -80°C.

## Chapter 3: Results

### 3.1 RNA polymerase II largest subunit Rpb1 lacks a canonical CTD in *Tetrahymena*

In most eukaryotes, the largest subunit of RNA polymerase II, Rpb1, contains a canonical carboxy-terminal domain consisting of Tyrosine<sup>1</sup>-Serine<sup>2</sup>-Proline<sup>3</sup>-Threonine<sup>4</sup>-Serine<sup>5</sup>-Proline<sup>6</sup>-Serine<sup>7</sup> (Y<sup>1</sup>S<sup>2</sup>P<sup>3</sup>T<sup>4</sup>S<sup>5</sup>P<sup>6</sup>S<sup>7</sup>) heptapeptide repeats which are necessary for generating the phosphoserine code required for the recruitment of many major regulatory proteins (Phatnani and Greenleaf, 2006). To elucidate the nature of the CTD of Rpb1 in *Tetrahymena*, I aligned the translated open reading frame (ORF) of the last 40 peptides of the H domain of *RPB1* onward from the budding yeast *S. cerevisiae* (S.c.), human (H.s.), and the ciliated protozoa *P. tetraurelia* (P.t.) and *T. thermophila* (T.t.) using Clustal Omega (EBI) (Figure 6). The resulting sequence alignment illustrates the divergence and lack of the canonical repeat structure in the CTD of *Tetrahymena* Rpb1.

In preparing the alignment, the H-domain and linker regions were included for reference, and the indicated CTD heptapeptide repeats were outlined and numbered based on the human Rpb1 sequence as described by Hsin and Manley (2012). The formatting of this alignment including its organization around the human CTD region was modeled after a similar sequence alignment illustrating the lack of conservation in the RNAPII CTD of the protist *Trypanosoma brucei* presented by Das and Bellofatto (2009). As such, the first heptapeptide repeat in this alignment does not necessarily correspond to such in other organisms. The sequence alignment data is inclusive to the last amino acid of the *Tetrahymena RPB1* sequence.

Although the primary amino acid sequence of the *Tetrahymena* Rpb1 CTD did not reveal a visible canonical heptapeptide repeat structure, I calculated the abundance of serine, threonine,

tyrosine, and proline in order to assess whether it had a similar composition to the CTD of yeast and human, and could perhaps harbour the capacity to perform functions comparable to a canonical CTD. Serine abundance was pertinent, as they are the primary phosphorylation targets of the Rpb1 CTD (Bartkowiak, 2011). Threonine and tyrosine are also phosphorylatable amino acids, and their phosphorylation on the CTD of Rpb1 has been linked with 3' end processing of histone mRNA (Hsin *et al.*, 2011) and transcriptional elongation (Mayer *et al.*, 2012), respectively. Proline abundance was included due to its role in peptidyl-proline bond isomerization involved in fine-tuning the CTD phosphorylation state through Ssu72-mediated dephosphorylation of Ser<sup>5</sup>P (Werner-Allen *et al.*, 2011). The sequences used for calculating amino acid abundance were derived from the "CTD" region in Figure 6, and included the first amino acid of the first repeat through to the end of each respective sequence (Appendix C.9).

Compared to *S. cerevisiae*, *T. thermophila* contains 24% less serines, 41% less threonines, 27% less tyrosines, and half the level of prolines. In relation to humans, *T. thermophila* contains 21% less serines, 54% less threonines, 36% less tyrosines, and 53% less prolines (Table 3; see Appendix C.9 for calculations). As a comparison, the differences in serine, threonine, tyrosine, and proline abundances in the Rpb1 CTD of *T. thermophila* compared to that of *P. tetraurelia* were calculated as 7.3%, 6.3%, 39.7%, and 5.2%, respectively. Although the *Tetrahymena* CTD is not enriched for serine, the position of available serines may permit sufficient phosphorylation. While approximately 67% of serines in the *Tetrahymena* CTD (26/39) are present in the 2<sup>nd</sup>, 5<sup>th</sup>, or 7<sup>th</sup> residue position, there is no observable regularity to suggest they encode a potential phosphorylation pattern.

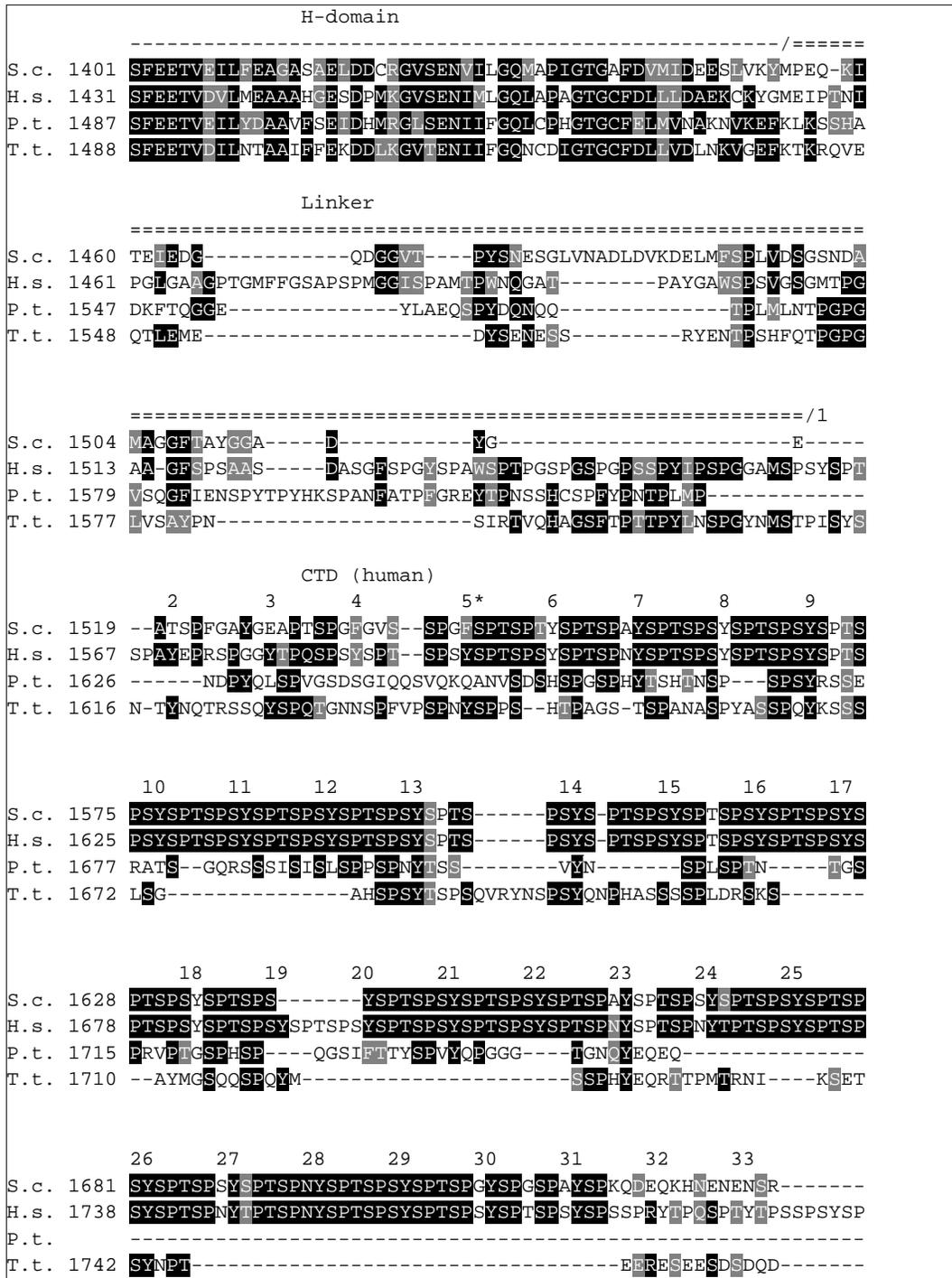


Figure 6. Multiple sequence alignment of the CTD of Rpb1

Alignment of the C-terminal region of Rpb1 encoded by human (H.s.), yeast (S.c.), *Paramecium* (P.t.) and *Tetrahymena* (T.t.) was achieved using Clustal Omega (EBI) and shaded with BoxShade 3.21 (ExPASy). The first 33 human heptapeptide repeats are labeled. "5\*" indicates the first yeast heptapeptide repeat. The linker region and last 40 residues of the H domain are included for reference. The alignment is inclusive of the last amino acid of *Tetrahymena* Rpb1.

Table 2. Percent abundance of serine, threonine, tyrosine, and proline in various Rpb1 CTDs

Percentage of serine and proline residues contained in the heptapeptide repeat region of the CTD of Rpb1 was compared between yeast, human, *Paramecium*, and *Tetrahymena*.

	Serine %	Threonine %	Tyrosine %	Proline %
Yeast (S.c.)	34.9 (75/215)	12.6 (27/215)	12.1 (26/215)	25.6 (55/215)
Human (H.s.)	33.6 (127/378)	16.1 (61/378)	13.8 (52/378)	27.5 (104/378)
<i>Paramecium</i> (P.t.)	24.6 (31/126)	7.9 (10/126)	6.3 (8/126)	13.5 (17/126)
<i>Tetrahymena</i> (T.t.)	26.4 (39/148)	7.4 (11/148)	8.8 (13/148)	12.8 (19/148)

### 3.2 Mediator subunit Med31 is conserved in eukaryotes

Subunit Med31 of the Mediator complex is the only Mediator subunit with homology in *Tetrahymena*. I generated a sequence alignment of the entire translated ORF of *MED31* from the budding yeast *S. cerevisiae*, fission yeast *S. pombe*, human, mouse, and *Tetrahymena* with Clustal Omega (EBI) to show coverage of sequence similarity over the entire gene (Figure 7). Protein BLAST analysis revealed 53% sequence identity between *Tetrahymena* and human Med31 covering 56% of the gene, and 33% sequence identity across 64% of the gene between that of *Tetrahymena* and yeast. In comparison, human and yeast Med31 share 44% identity with a coverage of 69%. When other Mediator subunits were queried against the *Tetrahymena* protein database using BLASTP, no identified proteins or "hits" were found. Such high sequence similarity is very interesting since Med31 is non-essential for cell viability in either yeast or *T. thermophila*, and other Mediator subunits that are less conserved tend to be essential for yeast viability.

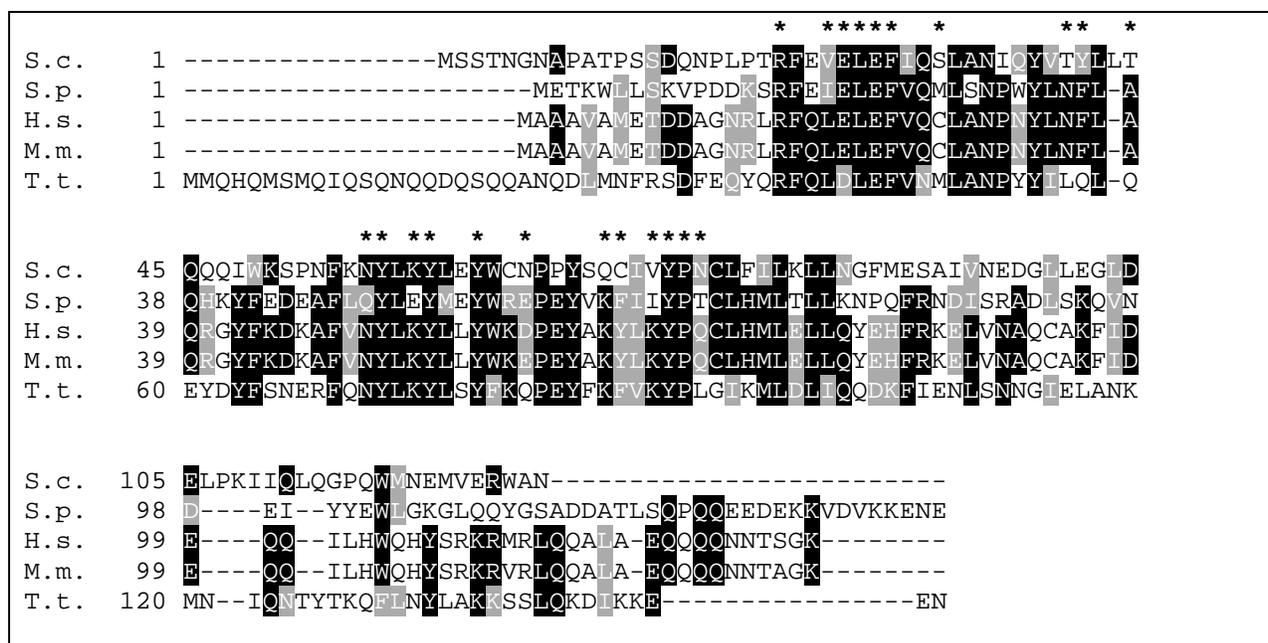


Figure 7. Multiple sequence alignment of Med31

Alignment of the entire translated ORF of *MED31* from budding yeast (*S.c.*), fission yeast (*S.p.*), human (*H.s.*), mouse (*M.m.*), and *Tetrahymena* (*T.t.*) by Clustal Omega. Shading by BoxShade 3.21. "\*" represents the Med31/Med7 interaction sites as described by Koschubs *et al.* (2009). E-values in relation to *S. cerevisiae* Med31 calculated by BLASTP against each organism's database: *S.p.*=5.5e-14, *M.m.*=3.8e-13, *H.s.*=5e-14, *T.t.*= 8e-07.

### 3.3 *MED31* of *T. thermophila* does not rescue the slow-growth phenotype of *S.*

#### *cerevisiae* deleted for *MED31*

The high sequence similarity between *MED31* of *S. cerevisiae* (*MED31*<sup>Sc</sup>) and *T. thermophila* (*MED31*<sup>Tt</sup>) prompted an investigation into whether the two proteins share similar attributes. To do this, I tested the ability of *Tetrahymena MED31* to rescue the slow-growth phenotype characteristic of yeast deleted for *MED31* (*med31*<sup>Sc</sup>Δ). Wildtype OB<sup>+</sup> strains (Open Biosystems) were used as the background for all transformations.

In order to synthesize a myc-tagged version of Med31, I cloned *MED31*<sup>Tt</sup> and *MED31*<sup>Sc</sup> into 12myc-pRb415 vectors with *Bam*HI and *Sal*I restriction enzymes. When cloned into this

vector, an N-terminal epitope tag is added to the target gene. Next, I transformed the constructs into a Med8-TAP yeast strain deleted for *MED31* (Med8-TAP/*med31*<sup>Sc</sup>Δ). Since the nature of this investigation was whether *Tetrahymena MED31* could substitute for yeast *MED31*, it was imperative that I started with a yeast strain deleted for *MED31* by which to directly compare the effect of adding Med31<sup>Tt</sup>. The Med8-TAP strain transformed with empty 12myc-pRb415 vectors acted as a control for wildtype growth. Likewise, the transformation of knockout Med8-TAP/*med31*<sup>Sc</sup>Δ strain transformed with *MED31*<sup>Sc</sup> was a control for phenotypic rescue.

The knockout Med8-TAP/*med31*<sup>Sc</sup>Δ strain transformed with the same empty 12myc-pRb415 vector was a negative control to ensure that any phenotypic rescue was a result of the *MED31* gene product and not the vector itself. Cloning was confirmed by sequencing with primer HJ559 (Appendix C.7).

The 12myc-pRb415 vector was not only used to impart an N-terminal 12xmyc epitope tag to target proteins, but because it is a CEN-based plasmid which contains a centromere sequence. In contrast to other vectors such as the high-copy 2μ plasmids, the centromere sequence bestows the ability for CEN-based vectors to replicate and segregate akin to small chromosomes and maintain a relatively low copy number close to WT. This is useful to mimic expression levels as if from a single gene locus. This bypasses the need for genomic yeast Med31 for homologous recombination, as it is imperative for genomic Med31 to have been disrupted. The 12myc-pRb415 vector also contains a yeast origin of replication for self-replication, a yeast *ADHI* promoter for the expression of 12xmyc-tagged genes, and a leucine marker (leu<sup>+</sup>) which enables yeast to grow on minimal media lacking leucine (YNB -leu). The selection of such leucine autotrophs on -leu media in turn identifies positive transformants.

Observations of growth following a 48 hour incubation at 30°C on minimal -leu media revealed that *MED31<sup>Tt</sup>* failed to restore wildtype growth rates as compared to *MED31<sup>Sc</sup>* when they were both transformed into yeast strains deleted for *MED31* (Figure 8). Med8-TAP/*med31<sup>Sc</sup>Δ* transformed with plasmid expressing 12myc-*MED31<sup>Sc</sup>* under the constitutive *ADHI* promoter displayed wildtype growth phenotype comparable to Med8-TAP transformed with empty 12myc-pRb415 vector, but when transformed with plasmid expressing 12myc-*MED31<sup>Tt</sup>*, wildtype growth was not rescued, and instead resembled the negative control Med8-TAP/*med31<sup>Sc</sup>Δ* transformed with empty 12myc-pRb415 vector (P).

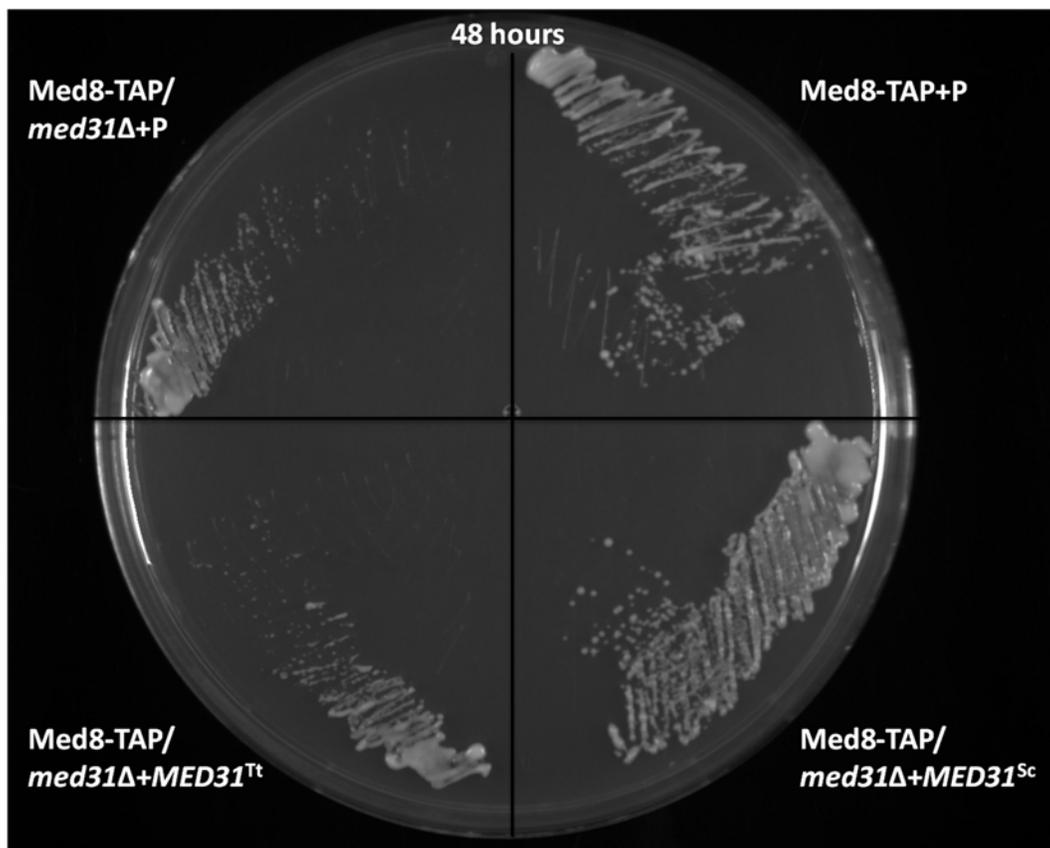


Figure 8. Growth of *med31<sup>Sc</sup>Δ* yeast transformed with yeast and *Tetrahymena MED31*

Colonies of Med8-TAP and Med8-TAP/*med31<sup>Sc</sup>Δ* were transformed with 12myc-pRb415, and Med8-TAP/*med31<sup>Sc</sup>Δ* was transformed with 12myc-*MED31<sup>Sc</sup>* and 12myc-*MED31<sup>Tt</sup>*. Cells were grown on -leu media for 48 hours at 30°C.

### 3.4 12myc-Med31<sup>Tt</sup> does not co-purify with Med8<sup>Sc</sup>-TAP

Although *Tetrahymena* Med31 failed to rescue the phenotype of the yeast *med31*<sup>Sc</sup> $\Delta$  deletion strain, this does not preclude its interaction with the yeast Mediator complex. To examine possible conservation of Mediator binding characteristics in *Tetrahymena* Med31, I tested its capacity to co-purify with yeast Mediator subunit Med8. Med8 was chosen for this analysis as it was shown to co-purify with many Mediator subunit peptides including Med31 in raw MS data provided by the Yeast TAP Project (<http://tap.med.utoronto.ca/>) and so was a prime candidate for determining whether it could interact with *Tetrahymena* Med31. The same strains used in the previously mentioned phenotypic rescue experiment were used for the co-purification study. Similar considerations were taken in using a yeast strain deleted for *MED31* (*med31*<sup>Sc</sup> $\Delta$ ) to prevent native yeast Med31<sup>Sc</sup> from sequestering Med8-TAP during affinity purification.

The input acts as a control to confirm all necessary protein constituents are present and expressed at the correct molecular weight corresponding to the protein plus epitope tag. These showed the presence of TAP and myc epitope tags in the starting material prior to affinity purification (Figure 9). Since all strains contained TAP-tagged Med8, it was expected that a signal representing Med8-TAP (~46kDa) would be present in all input samples as a control for proper expression of the Med8-TAP fusion protein and effective binding by the  $\alpha$ -TAP antibody, while only the two strains transformed with 12myc-Med31 would reveal probing by  $\alpha$ -myc antibody. Signals for myc were observed in strains transformed with 12myc-Med31<sup>Tt</sup> (32.6kDa) and 12myc-Med31<sup>Sc</sup> (29.1kDa) as expected. The visualization of the 12myc tag in the input acted as a loading control for later to signify that the absence of myc signal in affinity purified samples was due to inefficient binding between Med8 and Med31, and not the lack of expressed tagged protein.

One-step affinity purifications of Med8-TAP from Med8-TAP/*med31*<sup>Sc</sup>Δ/12myc-*MED31*<sup>Sc</sup>, Med8-TAP/*med31*<sup>Sc</sup>Δ/12myc-*MED31*<sup>Tt</sup>, OB<sup>+</sup>/Med8-TAP and Med8-TAP/*med31*<sup>Sc</sup>Δ transformed with empty 12myc-pRb415 vector revealed that yeast Med31 but not *Tetrahymena* Med31 can interact with Med8 of the *S. cerevisiae* Mediator complex (Figure 9). The presence of α-TAP signal in all affinity-purified samples was a control which indicated proper IgG binding and protein elution. In addition, the accompaniment of a signal for myc should only be present if Med8-TAP, being the bait bound to IgG, had successfully bound to the 12-myc-Med31 prey.

It is known that Med31 and Med8 interact with each other in *S. cerevisiae*, so the appearance of a signal for 12myc-Med31<sup>Sc</sup> following Med8-TAP affinity purification was a control for ensuring that the N-terminal 12xmyc and the C-terminal TAP tags had not interfered with Med8-Med31 binding by altering the conformation of the protein subunits or otherwise. Appropriate signal representing 12myc-Med31<sup>Sc</sup> but not 12myc-Med31<sup>Tt</sup> was observed in the affinity-purified samples and suggests that *Tetrahymena* Med31 cannot bind yeast Mediator.

Transformation of 12myc-Med31<sup>Sc</sup> acted as a positive control in that it represented effective co-purification with Med8-TAP by which to compare *Tetrahymena* Med31 affinity purification. To dismiss the possibility of IgG binding Med31<sup>Sc</sup> directly, affinity purification should be carried out on an untagged wildtype strain that has been transformed with 12myc-Med31<sup>Sc</sup>. Since 12myc-Med31 should not bind to IgG and would be washed away, no signal for myc would be expected. A positive signal would indicate 12myc-Med31<sup>Sc</sup> was purified by IgG without Med8-TAP as an intermediate and thus negate the validity of the study. Lack of slow-growth rescue by Med31<sup>Tt</sup> in *med31*<sup>Sc</sup>Δ or its interaction with Med8-TAP in yeast shows the inability of *Tetrahymena* *MED31* to complement yeast knocked out for *MED31*, and precluded further analysis in yeast.

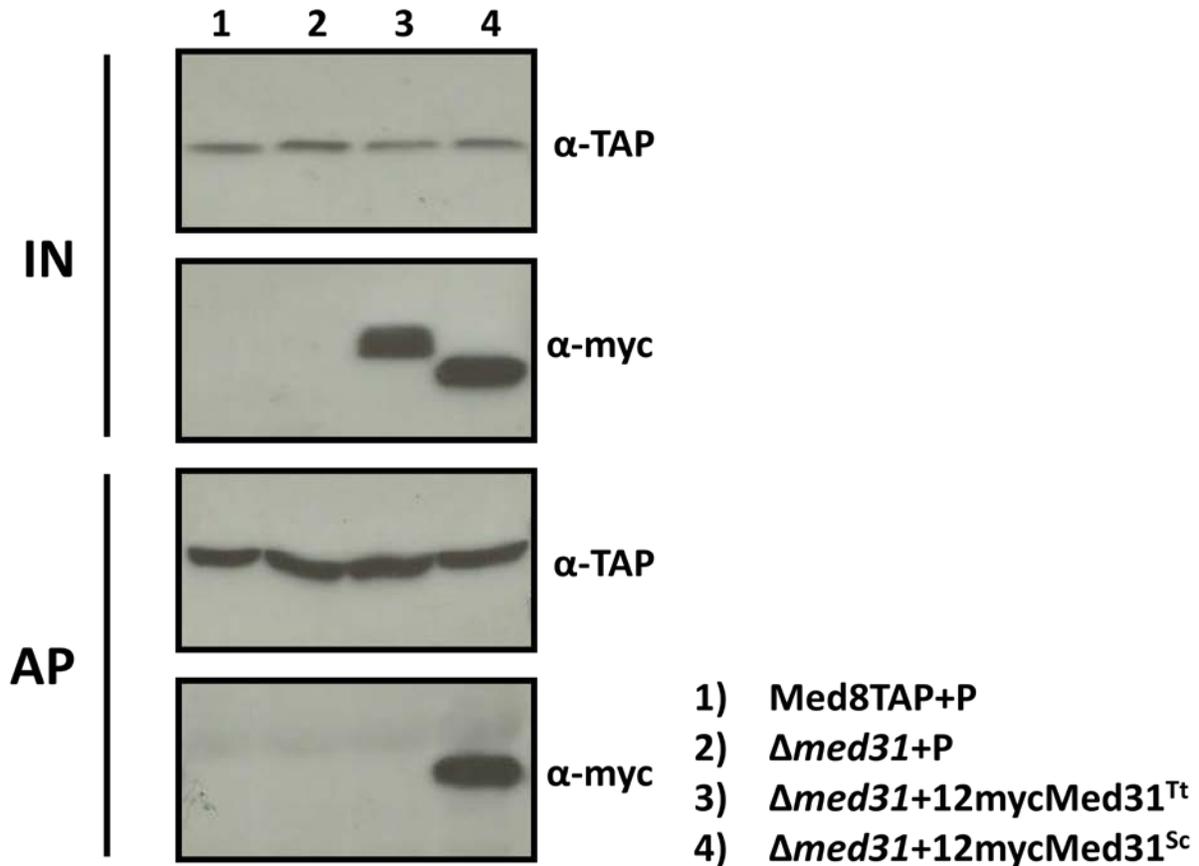


Figure 9. Western blot of Med8-TAP affinity purification

Med8-TAP affinity purifications were performed in strains  $OB^+$ /Med8-TAP +P, Med8-TAP/ $med31^{Sc}\Delta$  +P, Med8-TAP/ $med31^{Sc}\Delta$ /12myc-MED31<sup>Tt</sup>, and Med8-TAP/ $med31^{Sc}\Delta$ /12myc-MED31<sup>Sc</sup>. Samples were separated by SDS-PAGE, transferred to PVDF membrane, and probed with monoclonal  $\alpha$ -TAP (1:1000) and  $\alpha$ -myc (1:3000) antibodies. Film exposure=1minute.

### 3.5 Construction of 3xFLAG-TEV-ZZ tagging cassette

By using a cassette with an epitope tag fused to the resistance marker and carefully designing flanking regions with homology to the start or stop codon, an endogenous macronuclear gene can be tagged at the amino or carboxy terminus for downstream chromatographic or subcellular localization purposes. Of the 7 potential Mediator and 5 potential Integrator subunits previously identified by affinity purification-mass spectrometry of *T.*

*thermophila* Med31 and Dss1, respectively (Table 3; Fillingham, unpublished), I selected Med3, Med4, Med17, Med20, Med22, and all 5 of the Integrator subunits for epitope tagging in order to further characterize the *Tetrahymena* Mediator and Integrator complexes.

In order to affinity purify putative Mediator and Integrator subunits from *T. thermophila*, I tagged the corresponding protein-encoding gene with a C-terminal triple FLAG epitope tag (3xFLAG) fused to a tobacco etch virus (TEV) protease cleavage site, fused to two repeats of the Z domain (ZZ) of the *Staphylococcus aureus* protein A (3xFLAG-TEV-ZZ, or FZZ). To accomplish this, I generated a tagging cassette by cloning 5' and 3' homologous sequences into a pBluescript-FZZ vector (pBKS-FZZ). The tagging cassette is depicted along with relevant features in a schematic I created using Microsoft Paint and Adobe Photoshop CS3 (Figure 10). I similarly designed a graphic representation of the tagging strategy (Figure 11).

The two-step cloning of homologous DNA sequences into the vector ensured proper orientation of the tagging construct relative to the endogenous sequence to allow the introduction of the *neo* resistance gene into the macronuclear genome. This was necessary for proper expression of gene fusion products as well as proper transcription of the *neo* gene required for paromomycin resistance. The *neo* gene is not endogenous to *Tetrahymena*, and therefore does not compete with target genes or target an ectopic site for recombination. Once the construct was within the macronucleus, it recombined homologously with the endogenous locus through recognition of the 5' and 3' homology sequences of their genomic counterparts to achieve exact gene replacement with everything between the flanking sequences. In this way, wildtype levels of gene expression were maintained as transcription remained under the control of the native promoter. The active histone H41 promoter directs high levels of *neo* gene transcription to ensure adequate resistance to paromomycin.

In transforming the macronucleus with a tagging or knockout cassette containing an antibiotic resistance marker (*neo*), it was required that at least one wildtype copy of the gene be replaced with this cassette through homologous recombination so that selection and phenotypic assortment of transformed cells could be achieved through antibiotics. Cells were only viable in antibiotic media if they contained enough *neo* transcript to meet the minimum threshold demands imposed by the drug. Since one of the two daughter cells after cell division invariably contained more copies of the tagged gene linked to the paromomycin resistance *neo* gene, phenotypic assortment accelerated by paromomycin (Figure 3) caused subsequent fission products to contain increased copies of *neo* transcript and thus the associated tagged gene. By gradually increasing the concentration of antibiotic for which the transformed cell carried minimal resistance to, cells were forced to contain increasing numbers of resistant copies in each generation until the macronucleus was homozygous for paromomycin resistance.

Table 3. Mass spectrometry data for *Tetrahymena* Dss1-FZZ

Data from liquid chromatography-mass spectrometry (LC-MS) following affinity purification of *Tetrahymena* Dss1/Rpn15. Green=bait Dss1/Rpn15 peptide.

	Gene ID	# OF PEPTIDES	NAME
	TTHERM_00227230	7	Rpn15
	TTHERM_00339610	425	Rpn1
	TTHERM_00442210	337	Rpn2
	TTHERM_00446090	62	Rpn3
	TTHERM_00578940	58	Rpn5
	TTHERM_00378970	116	Rpn6
	TTHERM_00191240	32	Rpn7
	TTHERM_00267990	35	Rpn8
19S Proteasomal Proteins	TTHERM_00388440	46	Rpn9
	TTHERM_00471830	29	Rpn10
	TTHERM_00049450	9	Rpn11
	TTHERM_00649110	160	Rpn12
	TTHERM_00279670	100	Rpt1
	TTHERM_01014660	104	Rpt2
	TTHERM_00068110	113	Rpt3
	TTHERM_00469100	137	Rpt4
	TTHERM_00136360	116	Rpt5
	TTHERM_00551090	105	Rpt6
19S Regulatory Particle Chaperones	TTHERM_00188860	33	Nas2/p27
	TTHERM_00954210	13	Hsm3/S5b
	TTHERM_00040320	13	Txn11
20S Core Subunit	TTHERM_00106960	4	Sc11
Ubiquitin	TTHERM_00339620	9	Ribosomal fusion protein
<i>BRCA2</i>	TTHERM_00437260	4	Brca2
Integrator	TTHERM_00532780	4	Int2
	TTHERM_00467840	13	Int4
	TTHERM_01243440	6	IntS6
	TTHERM_01159920	8	Int9
	TTHERM_00339790	5	Int11

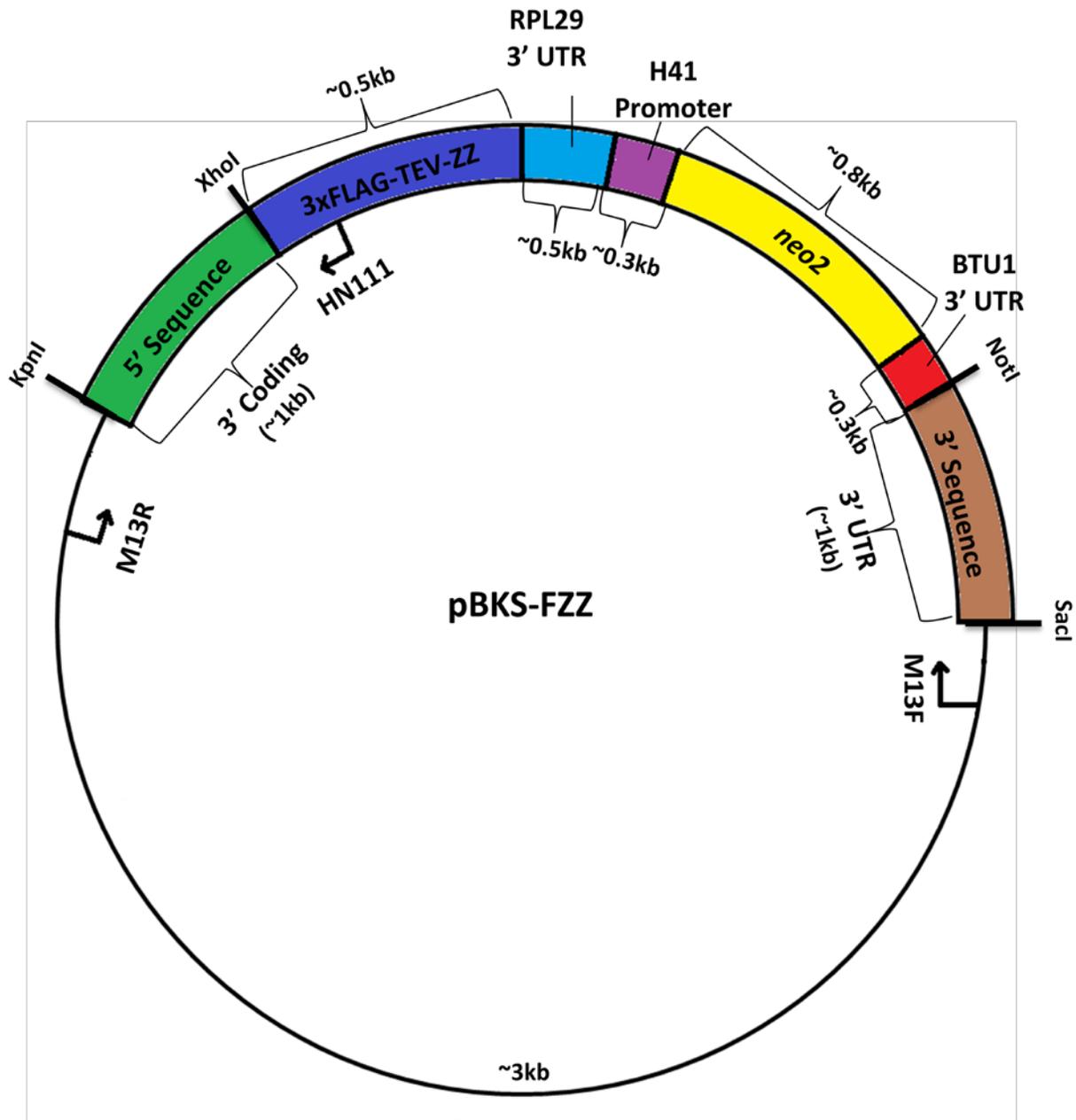


Figure 10. Schematic of *Tetrahymena* FZZ-tagging vector pBKS-FZZ

The pBKS-FZZ vector was designed to add a DNA sequence encoding a C-terminal FZZ epitope tag to protein-encoding genes in *T. thermophila*, selectable through its *neo* resistance gene. Schematic designed with Microsoft Paint and Adobe Photoshop CS3.

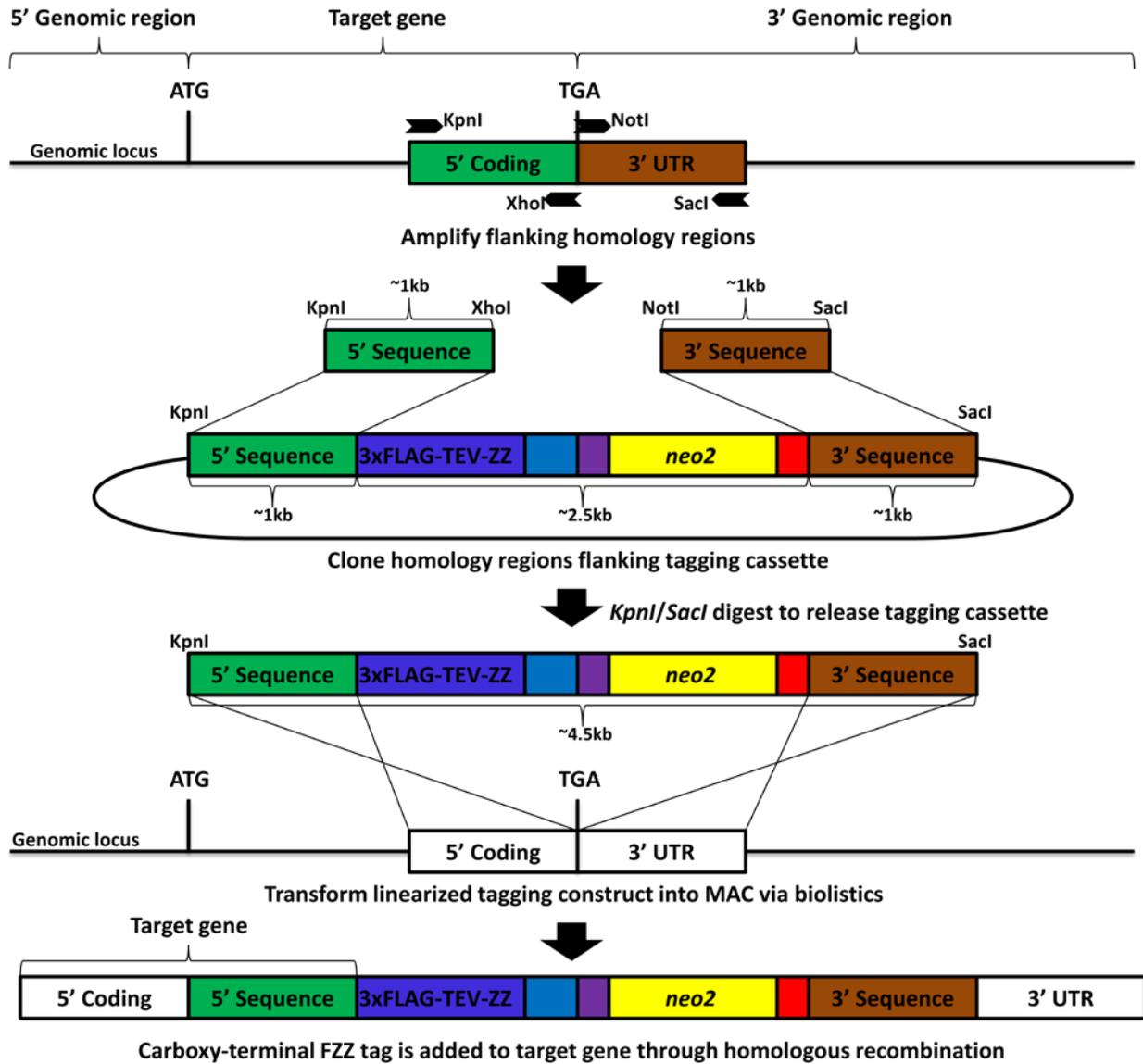


Figure 11. Schematic of protein-tagging strategy in *Tetrahymena*

Representation of the two-step directional cloning strategy used for generating FZZ-tagging cassettes to tag putative *T. thermophila* Mediator and Integrator subunits. Illustration designed with Microsoft Paint and Adobe Photoshop CS3.

### 3.5.1 PCR amplification of 5' and 3' homology sequences

The assembly of the C-terminal FZZ-tagging cassette for a particular gene involved two-step cloning to ensure collinearity of the homologous sequences in the tagging cassette with the

genomic locus for the homologous recombination in *Tetrahymena*. Thus, for each target gene it was necessary to PCR amplify 1kb sequences upstream and downstream of the stop codon, which ensures that translation after integration continues unhindered from the target gene through the FZZ tag to create a continuous gene fusion product. All PCR products were first separated on an agarose gel to ensure proper size and adequate concentration for downstream manipulation (Figure 12). The 3' sequence for *MED17* was deliberately truncated upstream of a *SacI* restriction site as it would be cleaved at this site during cassette linearization anyway (Fig 9), but otherwise, all homology regions were designed to be approximately 1kb to ensure sufficient sequence homology for integration of DNA into the macronucleus.

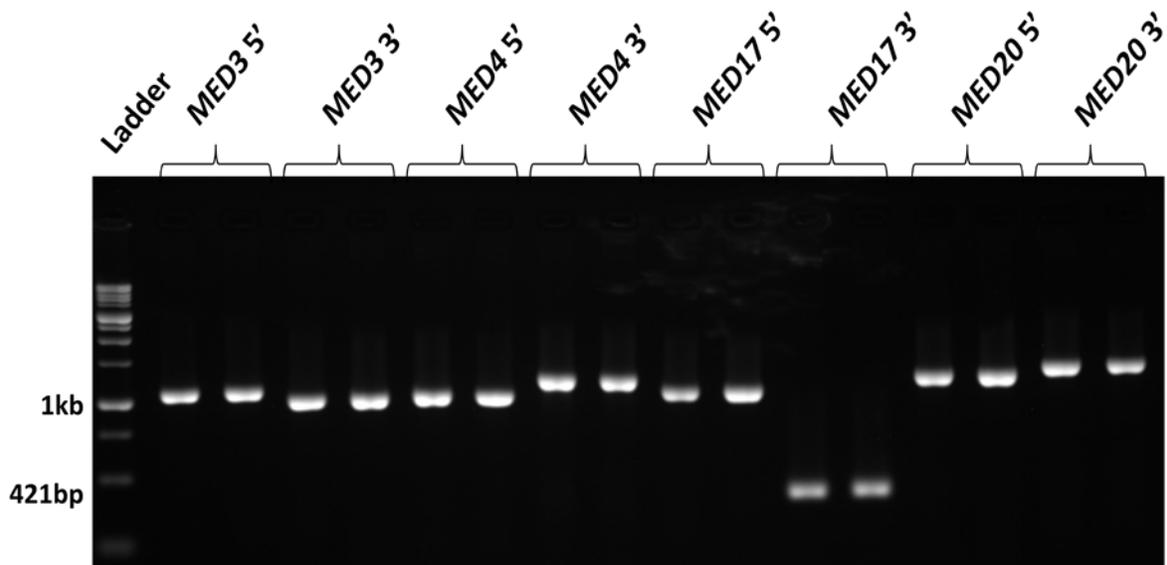


Figure 12. PCR products of 5' and 3' Mediator subunit homology sequences

As a representation of all 5' and 3' FZZ-tagging cassettes, 1 $\mu$ l of 5' and 3' PCR products of *Tetrahymena* Mediator subunits *MED3*, *MED4*, *MED17*, and *MED20* were electrophoresed on a 1% agarose gel at 80V. 5 $\mu$ l of DNA ladder was loaded.

### 3.5.2 Diagnostic restriction digest

After 5' and 3' homology sequences were cloned into the pBKS-FZZ vector, small-scale diagnostic restriction digests were performed to verify their correct integration. The samples were electrophoresed on an agarose gel (Figure 13), and only after confirmation that the released fragments were of the correct size (~1kb) were the plasmids submitted for sequencing.

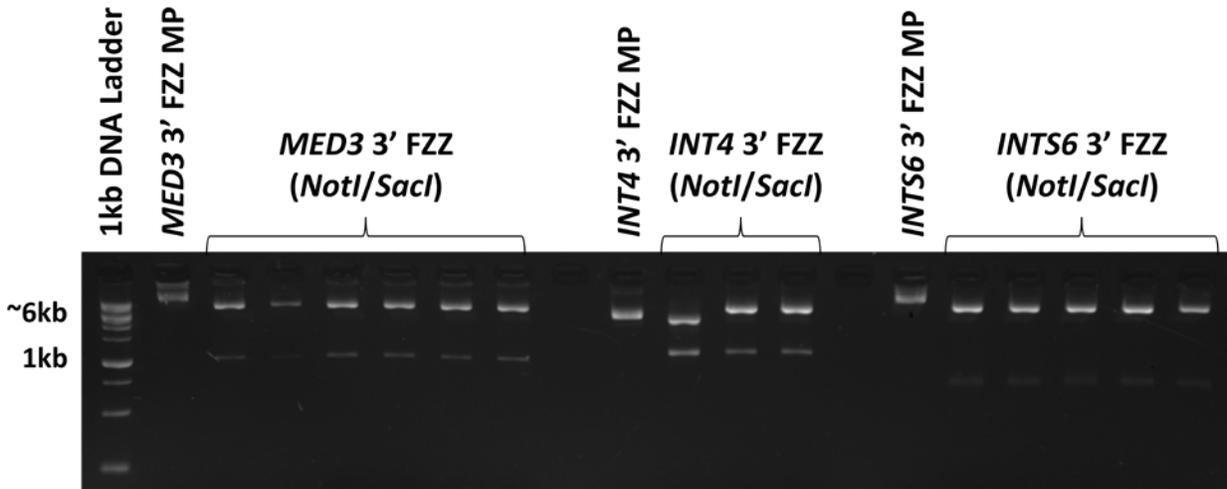


Figure 13. Diagnostic restriction digest of representative FZZ tagging vectors

pBKS-FZZ with cloned 3' sequences of *MED3*, *INT4*, and *INTS6* were digested with *NotI* and *SacI*. 1 $\mu$ l of each sample and 5 $\mu$ l of 1kb DNA ladder were electrophoresed on a 1% agarose gel at 80V.

### 3.5.3 Confirmation of FZZ-tagging cassettes

Sequencing of full FZZ-tagging vectors after the cloning of 5' and 3' homology sequences using primers M13R or HN111 (Figure 10; Appendix C.7) for confirmation of 5' sequence and primer M13F for confirmation of the 3' sequence was achieved for *MED3*, *MED20*, *MED22*, *INT4*, *INTS6*, and *INT11*.

### 3.5.4 Linearization of tagging construct for transformation into *T. thermophila*

Homologous recombination by exact gene replacement required that the construct be linearized prior to transformation. Complete pBKS-FZZ vectors with both 5' and 3' homology sequences confirmed by DNA sequencing were linearized with *KpnI* and *SacI* restriction endonucleases to release the 4.5kb tagging cassette from the vector. The resulting digestion products were electrophoresed on a 1% agarose gel for confirmation of correct number and size of fragments prior to transformation into the *Tetrahymena* macronucleus via biolistics (see Materials and Methods). Upon linearization, samples including Int4-FZZ, IntS6-FZZ, Med4-FZZ, and Med20-FZZ presented a 4.5kb signal representing the tagging cassette as well as a 3kb signal indicative of the resulting plasmid "backbone" (Figure 14).

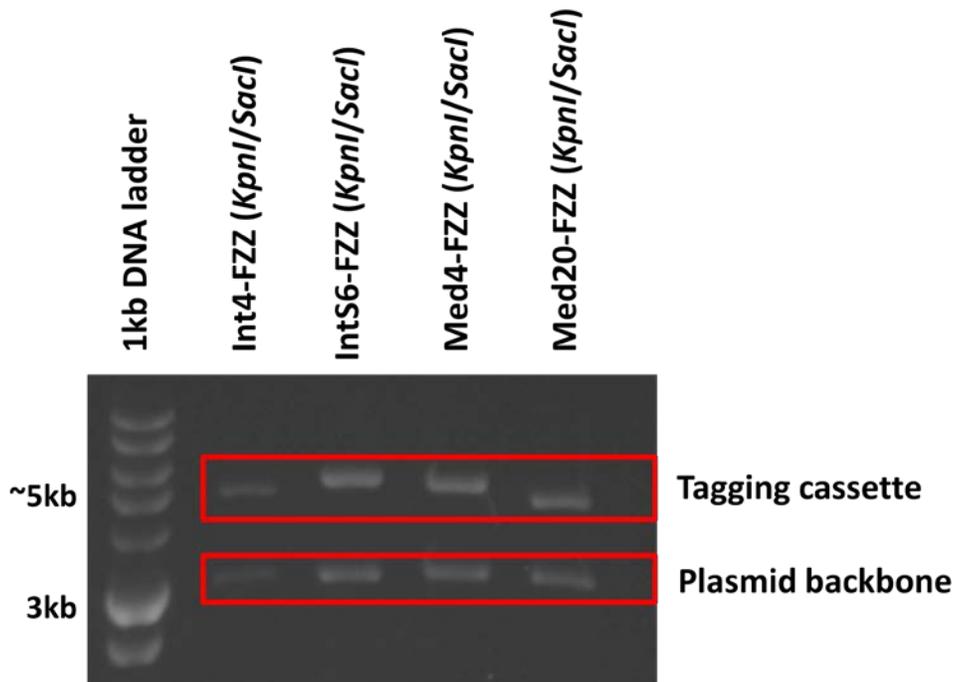


Figure 14. Gel electrophoresis of representative linearized tagging construct

Complete pBKS-FZZ vectors for Int4, IntS6, Med4, and Med20 were digested with *KpnI* and *SacI* to release linearized tagging cassettes. 1 $\mu$ l was each sample and 5 $\mu$ l of 1kb DNA ladder were separated on a 1% agarose gel at 80V.

### **3.6 Int4-FZZ and Med20-FZZ screening Western blots**

Western blots of TCA-extracted Int4-FZZ and Med20-FZZ were run to screen transformants for FZZ tagging (Figure 15). Due to the large molecular weight of Med20-FZZ (134kDa) and Int4-FZZ (145kDa), samples were run on 5% gels at 80V with a higher range protein ladder, Spectra Multicolor Broad Range Protein Ladder (Thermo Scientific; Appendix C.6), range 10-260kDa, until all fragments below 70kDa were run off the gel. After probing with primary  $\alpha$ -FLAG (Sigma-Aldrich) and  $\alpha$ -actin (Cedarlane) antibodies (see Materials and Methods), both Med20-FZZ and Int4-FZZ blots probed with  $\alpha$ -FLAG revealed a strong non-specific band slightly larger than 100kDa in all lanes including wildtype. Although appropriate sized signals against  $\alpha$ -FLAG and  $\alpha$ -actin were seen for both Med20-FZZ and Int4-FZZ, a signal against  $\alpha$ -FLAG around 140kDa was apparent in the wildtype lane despite using another source for fresh wildtype culture and remaking protein extracts. This obstructed the validity of the presence of the tag, and because a trustworthy blot was not achieved, Med20 and Int4 were not submitted for mass spectrometry.

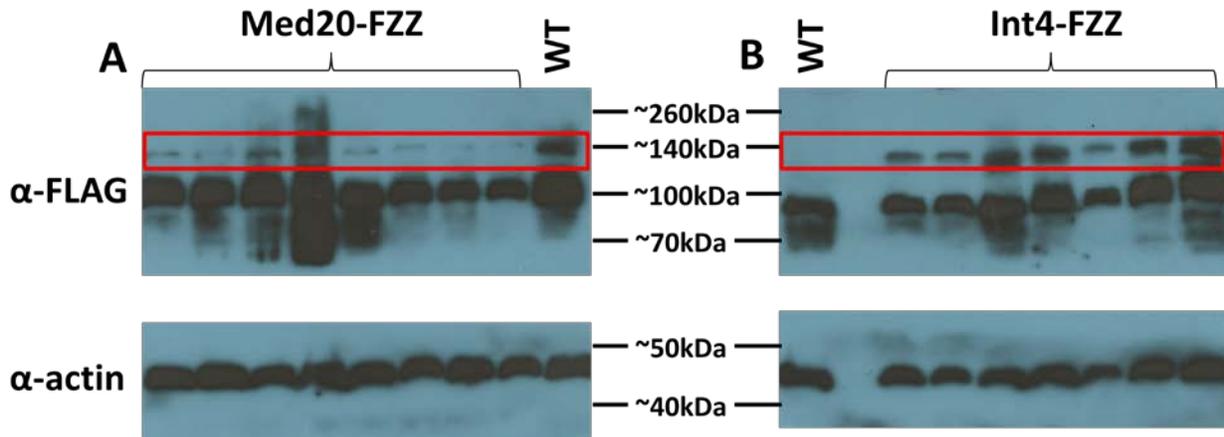


Figure 15. Western blot of Med20-FZZ and Int4-FZZ screening

A) Int4-FZZ screening of transformants reveals a band around 140kDa. A non-specific band near 100kDa is apparent in all lanes including wildtype. Signal for  $\alpha$ -actin was around 43kDa. B) Med20-FZZ screening of transformants reveals both 134kDa and 100kDa signal in all lanes. Signal for  $\alpha$ -actin was around 43kDa. All samples were run on the same 5% SDS-polyacrylamide gel and transferred to a PVDF membrane, which was cut into the resulting four blots. Probing was performed with monoclonal  $\alpha$ -FLAG (1:500) and  $\alpha$ -actin (1:1000) antibodies. Film exposure=5 minutes.

### 3.7 Confirmation of FZZ tagging of *Tetrahymena* Med22

Following phenotypic assortment to generate homozygous Med22-FZZ strains, whole cell extracts (WCEs) were generated from Med22-FZZ using TCA precipitation. Western blot analysis of these WCEs using  $\alpha$ -FLAG (1:500 dilution) and  $\alpha$ -actin (1:1000 dilution) mouse monoclonal antibodies was performed to confirm the FZZ-tagging of Med22 in *Tetrahymena* (Figure 16). There was clear signal for FLAG around the predicted 41kDa in Med22-FZZ but not wildtype (WT) lanes probed with  $\alpha$ -FLAG antibody, indicating the bands for Med22-FZZ were not aberrant or non-specific, and representative of a 3xFLAG tag. The signal for actin was apparent around 43kDa in all lanes including wildtype probed with  $\alpha$ -actin antibody as expected, and was included as a loading control to signify that the vacancy of FLAG signal in wildtype sample was not due to absence of sample.

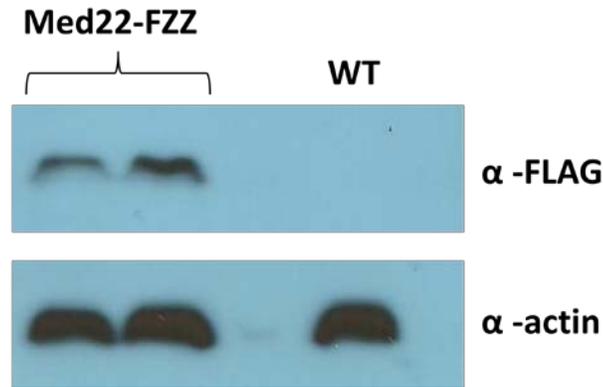


Figure 16. Western blot analysis of TCA-extracted putative *Tetrahymena* Med22-FZZ

TCA extracts were run in a 10% SDS-polyacrylamide gel with a 5% stacking gel at 100V, transferred to PVDF membrane and probed with monoclonal  $\alpha$ -FLAG (1:500) and  $\alpha$ -actin (1:1000) antibodies.

### 3.8 Med22 affinity purification-mass spectrometry in *Tetrahymena*

Three affinity purifications of Med22-FZZ were carried out, followed by Western blot analysis using  $\alpha$ -FLAG primary antibody (Sigma-Aldrich) at a 1:5000 dilution (see Materials and Methods). Correct size differentiation between input (IN) Med22-FZZ (~41kDa) and affinity purified Med22-3xFLAG following TEV cleavage and loss of the ZZ domain (~26kDa), as well as enrichment of signal intensity in immunoprecipitated Med22, suggests that Med22 had been successfully affinity purified (Figure 17). As there is not an associated blot probed with  $\alpha$ -actin to indicate whether wildtype WCEs contained protein, the absence of a signal in the wildtype immunoprecipitation (IP) does not contribute to the validity of the Med22 IP. The Med22 IP lane also appears to contain IgG heavy and light chains (53kDa and 23kDa, respectively). Samples from three affinity purifications of Med22-FZZ in *Tetrahymena* were submitted for analysis using tandem mass spectrometry (see Materials and Methods). In total, three peptides of the bait Med22 were identified, however, no other significant peptides were recognized (Table 4).

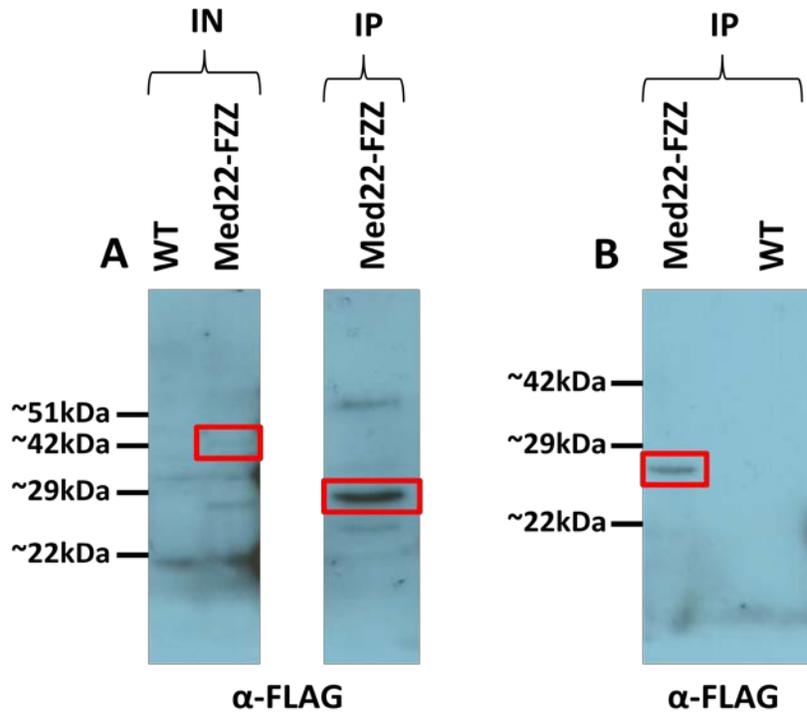


Figure 17. Western blot analysis of putative *Tetrahymena* Med22 affinity purification

Western blots from two separate Med22 affinity purifications. A) Jan.22/2013 purification, monoclonal  $\alpha$ -FLAG antibody (1:5000), film exposure=1 minute, B) Feb. 8/2013 purification, monoclonal  $\alpha$ -FLAG antibody (1:5000), film exposure=5 minutes.

Table 4. Tandem mass spectrometry data for *Tetrahymena* Med22-FZZ

The cumulative results of three sets of Med22-FZZ affinity purification-tandem mass spectrometry only identified three bait Med22 peptides. Green=bait Med22 peptide.

GENE ID	SPECTRAL COUNT	MED SUBUNIT
TTHERM_00670380	3	Med22

### 3.9 Med31 affinity purification-mass spectrometry in *Tetrahymena*

To confirm and build on previous *Tetrahymena* Med31 affinity purification data (Fillingham, unpublished), I repeated tandem affinity purification in *T. thermophila* with the previously constructed strain of *T. thermophila* expressing Med31-FZZ. Whole cell extracts from my affinity purifications were subjected to Western blot analysis using  $\alpha$ -FLAG primary antibody (Sigma-Aldrich) at a 1:5000 dilution and  $\alpha$ -actin primary antibody (Cedarlane) also at 1:5000 dilution. Following exposure onto film, observed signals were close to the predicted 36kDa for input Med31-FZZ and 21kDa for Med31-3xFLAG following TEV cleavage and loss of the ZZ domain of the tag (Figure 18). Compounded with the clean  $\alpha$ -actin blot revealing proper loading of all input samples and subsequent loss of actin through purification, the Western blot suggests Med31 was successfully affinity purified. The lower ~23kDa signal in the Med31-FZZ  $\alpha$ -FLAG input lane is most likely a non-specific protease digestion product as a fragment half the size of the parent signal is sometimes observed.

Cumulative mass spectrometry data from three sets of Med31 affinity purifications in *Tetrahymena* (Table 5) identified approximately 24 novel co-purifying proteins. Sequence analysis using BLASTP indicates that they do not share sequence similarity to anything in GenBank (NCBI), however, Position Specific Iterative (PSI)-BLAST-based bioinformatic analysis of Mediator in a variety of protists revealed that 8 of these 24 proteins are likely bona fide *Tetrahymena* Mediator orthologs (Bourbon, 2008). Also predicted by Bourbon (2008), these results revealed co-purification of putative *Tetrahymena* Med3; the first predicted Mediator subunit characterized in *Tetrahymena* belonging to the more divergent ancient tail module. Predicted *Tetrahymena* Mediator subunits 4, 7, 11, 17, 20, 21, and 22, which previously purified with *Tetrahymena* Med31-FZZ in an earlier pilot experiment (Fillingham, unpublished) were

again recovered with Med31-FZZ, reinforcing the composition of the *Tetrahymena* Med31 interactome. The "Gene ID" refers to the *Tetrahymena* accession number, while "spectral count" represents the number of total peptide spectra identified for a given peptide. Confidence scores of protein interactions are determined by significance analysis of interactome (SAINT) tool and given a probability of interacting with the bait ranging from 0 to 1; 0 signifying non-significance and 1 advocating statistical significance, and the peptides were considered significant if they met a probability cutoff value of 0.8. Peptides in Table 5 all had a SAINT score of 1 and therefore average probability was 1.

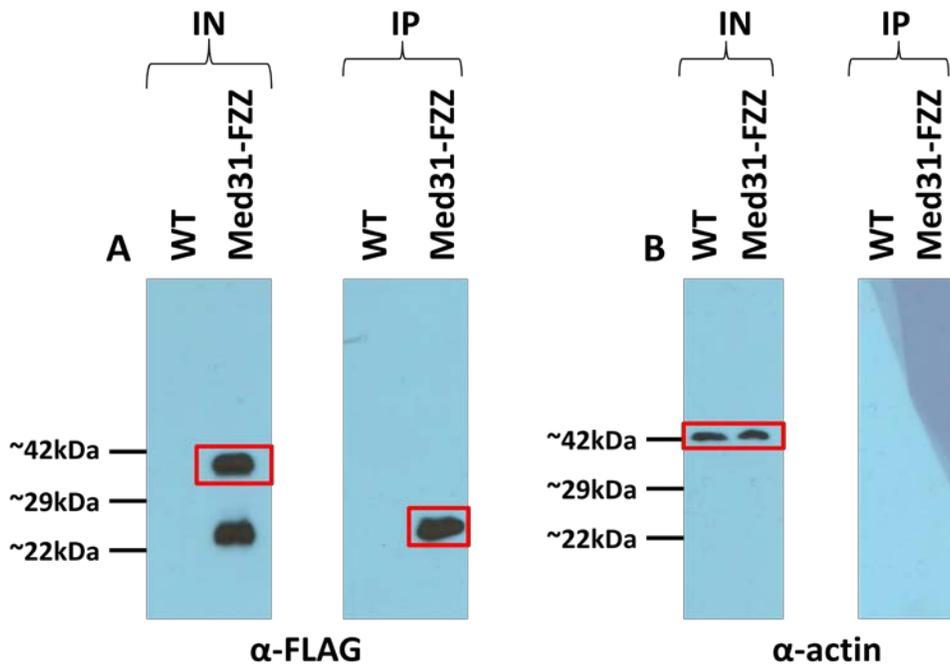


Figure 18. Western blot analysis of *Tetrahymena* Med31 affinity purification

*Tetrahymena* Med31 affinity purification Western blots. A) Probing with monoclonal α-FLAG antibody (1:5000), film exposure=30 seconds, B) Probing with monoclonal α-actin antibody (1:5000), film exposure=30 seconds.

Table 5. SAINT analysis of *Tetrahymena* Med31-FZZ affinity purification

Spectral count and SAINT analysis following three Med31-FZZ affinity purifications in *Tetrahymena*. Green=bait Med31 peptide, yellow=peptides previously identified as co-purifying with *Tetrahymena* Med31, orange= novel Med31 co-purifying peptide, red=potential *Tetrahymena* Mediator subunit identified in this study as predicted by Bourbon (2008).

GENE ID	SPECTRAL COUNT	SAINT	MED SUBUNIT
TTHERM_00355460	51 56 27	1	MED31
TTHERM_00467799	20 29 9	1	
TTHERM_00691210	33 44 42	1	MED4
TTHERM_00028490	28 46 25	1	
TTHERM_00334350	27 39 24	1	
TTHERM_00918460	20 23 18	1	MED11
TTHERM_00829330	25 9 9	1	
TTHERM_00732830	34 7 15	1	
TTHERM_00490640	37 19 14	1	
TTHERM_00444720	17 19 10	1	
TTHERM_01002760	60 20 34	1	
TTHERM_00295380	44 61 41	1	MED21
TTHERM_00316620	57 44 55	1	
TTHERM_00752180	74 97 60	1	
TTHERM_00989470	64 36 30	1	
TTHERM_00052189	64 35 63	1	
TTHERM_01014520	89 40 66	1	
TTHERM_00780600	128 78 79	1	MED17
TTHERM_00147570	74 63 58	1	MED7
TTHERM_00670380	74 74 41	1	MED22
TTHERM_00922930	75 40 36	1	MED20
TTHERM_00419920	52 27 36	1	
TTHERM_00657539	79 34 41	1	
TTHERM_00490630	129 44 58	1	MED3
TTHERM_00052180	187 93 135	1	

### 3.10 Int11-FZZ affinity purification-mass spectrometry in *Tetrahymena*

Western blots performed on Int11-FZZ affinity purified material (Figure 19) were subjected to probing with primary  $\alpha$ -FLAG (Sigma-Aldrich) and  $\alpha$ -actin (Cedarlane) antibodies, both at a dilution of 1:5000. Although the signal intensity for affinity purified Int11-3xFLAG was not enriched over its input, the isolation of a single, clean band in the vicinity of the expected 82kDa for Int11 IP probed with  $\alpha$ -FLAG was grounds for proceeding with mass spectrometry. Unfortunately, tandem mass spectrometry analysis of affinity purified Int11 identified no significant prey peptides or Int11 bait peptides, so affinity purification should be repeated.

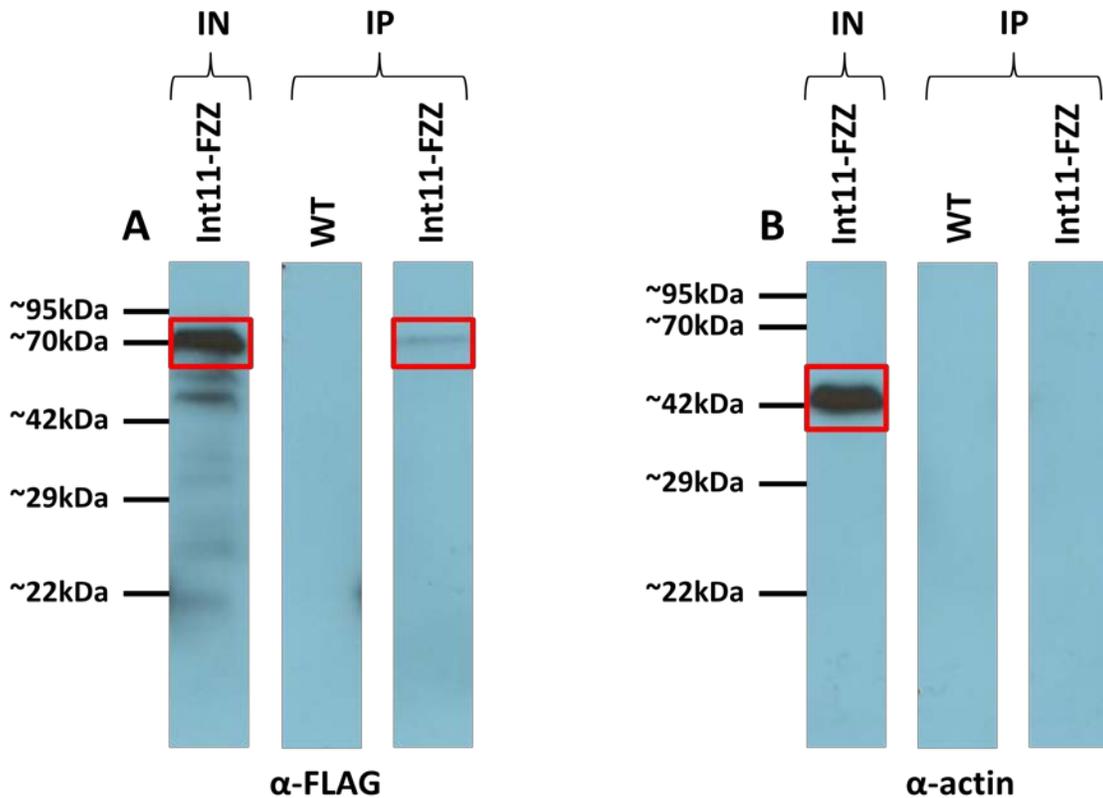


Figure 19. Western blot analysis of *Tetrahymena* Int11 affinity purification

A) Western blot probed with monoclonal  $\alpha$ -FLAG antibody (1:5000), film exposure=30 seconds.

B) Western blot probed with monoclonal  $\alpha$ -actin antibody (1:5000), film exposure=30 seconds.

### 3.11 Construction of the *INTS6* knockout cassette

By flanking a *neo* resistance gene with loci homologous to the 5' and 3' UTRs of a target gene, all macronuclear copies of a gene can be knocked out through exact gene replacement by homologous recombination, provided that the gene is non-essential for viability. Of the 5 potential Integrator subunits previously identified in *T. thermophila* by affinity purification-mass spectrometry of Dss1<sup>Tt</sup> (Fillingham, unpublished), I selected putative *Tetrahymena* Integrator subunit IntS6 for knocking out. Transformation led to four potential transformant colonies, but phenotypic assortment was halted at 200µg/ml paromomycin as cells did not flourish above this limit and died out at 500µg/ml paromomycin. As a comparison, the FZZ-tagged lines flourished in paromomycin concentrations up to 1200µg/ml.

In order to generate a *Tetrahymena INTS6* macronuclear knockout, directional cloning of 5' and 3' homology sequences flanking a *neoI* containing cassette was required. I designed two sets of primers specifically intended to amplify 1kb regions immediately upstream of the start codon and immediately downstream of the stop codon of *INTS6* to replace the entire *INTS6* gene with a *neoI* resistance gene through homologous recombination. A schematic I created of p4T2-1 and its relevant features (Figure 20) as well as a graphic representation of the knockout strategy (Figure 21) using Microsoft Paint and Adobe Photoshop CS3 provide a simplified overview of my knockout approach. Accurate two-step cloning of these *INTS6*-flanking sequences into the knockout vector p4T2-1 was confirmed by diagnostic digest (Figure 22) and subsequent sequencing using H4neo reverse (H4neoR) and BTU2R sequencing primers (Appendix C.7).

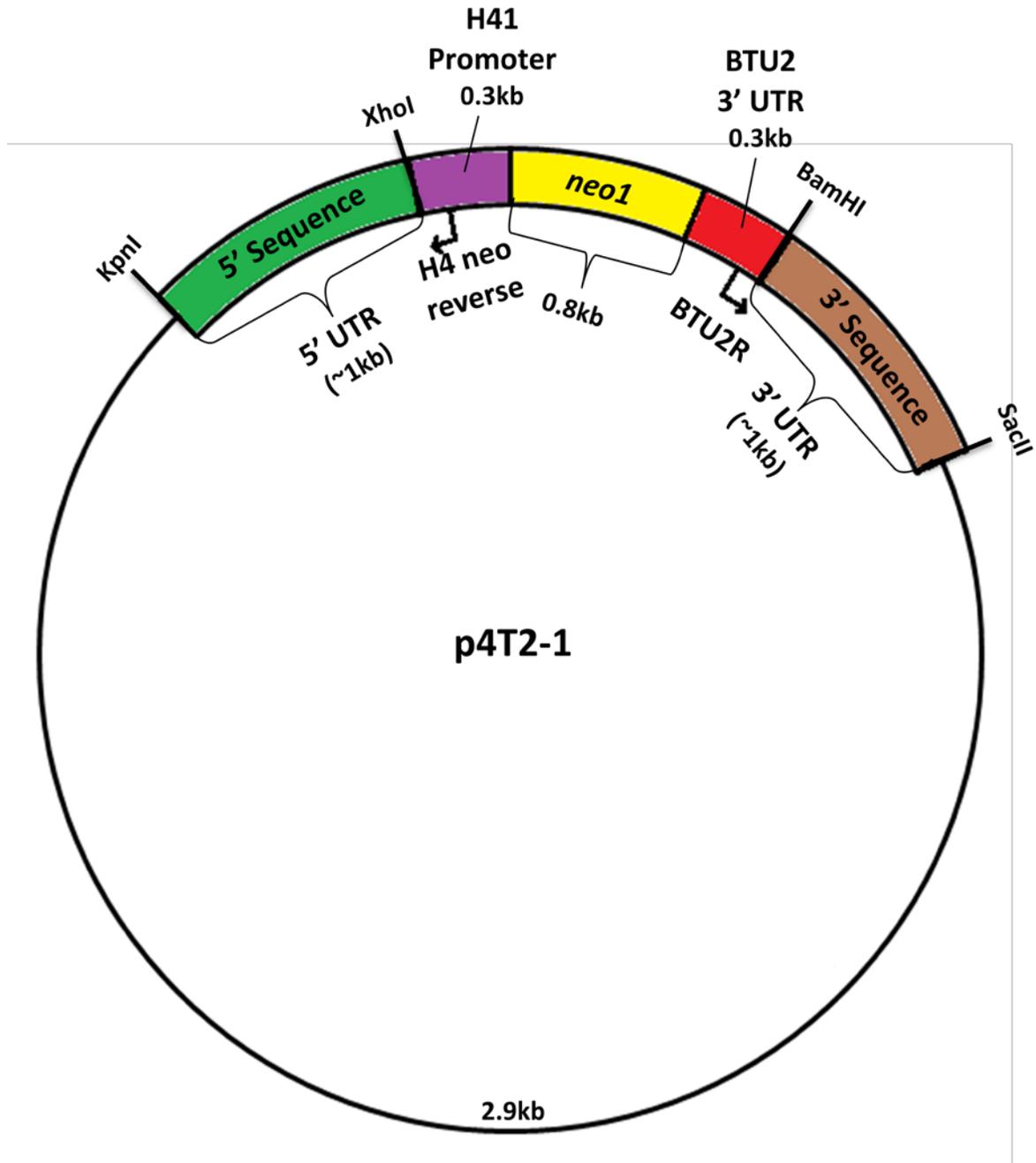


Figure 20. Schematic of *Tetrahymena* knockout vector p4T2-1

The p4T2-1 vector was used to replace protein-encoding genes in *T. thermophila* with a *neo1* cassette. Schematic designed with Microsoft Paint and Adobe Photoshop CS3.

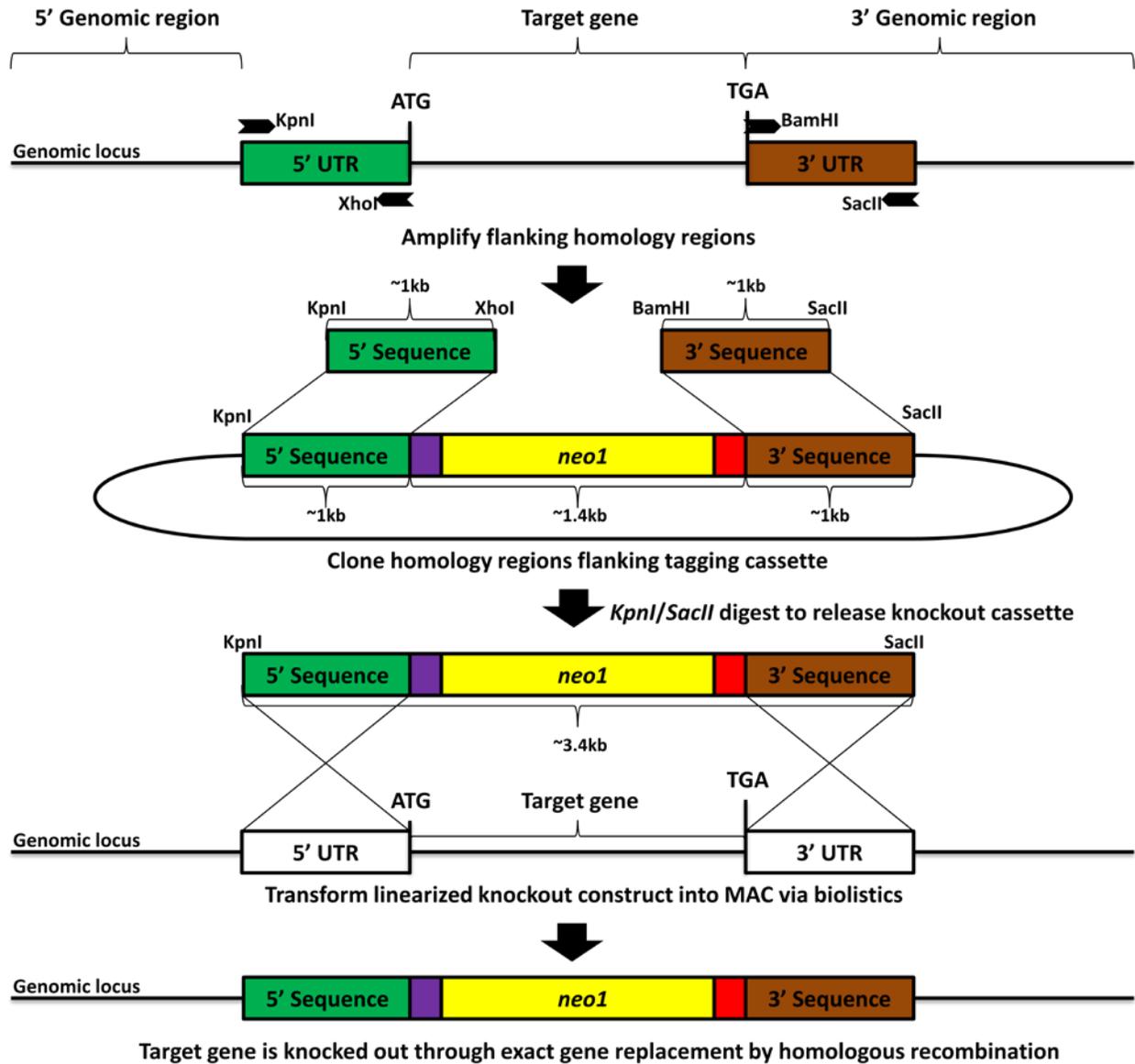


Figure 21. Schematic of *INTS6* knockout strategy in *Tetrahymena*

Representation of the directional cloning strategy used for knocking out putative *T. thermophila* Mediator and Integrator subunits. Illustration designed with Microsoft Paint and Adobe Photoshop CS3.

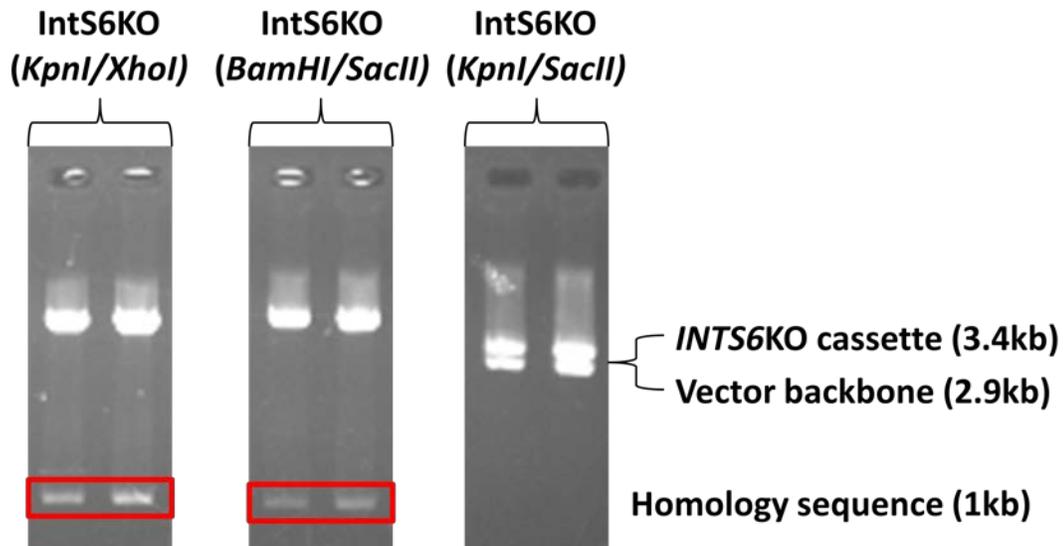


Figure 22. *INTS6*KO diagnostic digests

*INTS6*KO cassette digested with *KpnI/XhoI*, *BamHI/SacII*, and *KpnI/SacII* to verify correct integration of 5' and 3' homology sequences and cassette linearization. Products were electrophoresed on a 1% agarose gel at 80V.

### 3.11.1 Successful knockout of *INTS6* (colony PCR)

To determine whether the knockout cassette had successfully replaced the *INTS6* locus after homologous recombination, colony PCR was performed using 3 sets of primers. The first pair was composed of the same forward and reverse synthesis primers KpnIF and SacIIR originally used to create the homology sequence (Appendix C.8). These primers are homologous to the 5' and 3' UTRs of *INTS6* found in the knockout cassette and the endogenous *INTS6* gene locus, and therefore have the benefit of being able to report the presence of both the knockout cassette and the *INTS6* gene (Figure 23). Since both transformed and wildtype samples were expected to produce a PCR product using these primers, distinguishing successful knockouts from wildtype was achieved by the size of the resulting PCR product. When run on a 1% agarose gel, primers bound the endogenous gene in the wildtype strain resulting in a signal that was close

to the predicted size of 5.4kb, while each of the four knockouts displayed a signal of approximately 3.5kb indicative of the knockout cassette as well as a faint signal around 5kb suggestive of endogenous *INTS6* (Figure 24). The 5kb signal in the transformed cultures indicates the presence of full-length *INTS6*, which could be due to incomplete macronuclear knockout of *INTS6* or perhaps micronuclear *INTS6*. A flaw in this experimental design exists in that the KpnIF/SacIIR primer pair is specific to the knockout cassette and thus only reports the presence of the cassette within the cell and not its correct integration into the *INTS6* locus. This means that a smaller fragment indicative of the knockout cassette (3.4kb) would have been generated in all transformed strains regardless of where the cassette had correctly integrated, and was not completely indicative of an *INTS6* knockout.

The second pair of primers consisted of KpnIF paired with H4neoR (Appendix C.7), a reverse primer that bound the 5' *HHF1* promoter region of the knockout cassette (Figure 25, A). Unlike the KpnIF/SacIIR primer pair that displayed a signal for wildtype strains, the intent behind the KpnIF/H4neoR primer pair was to specifically identify only transformed cells and reveal no signal for wildtype or untransformed strains since the H4neoR primer is specific to the knockout cassette. While this could verify the presence of the knockout cassette within the cell, again it did not establish correct integration into the *INTS6* locus as these primers would bind all instances of the cassette.

The key to establishing correct integration into the *INTS6* locus rather than simply testing for the presence of the knockout cassette was to use a third pair of primers that coordinated in creating a PCR product that reported on the fusion of the endogenous locus with the cassette. I designed a forward primer designated "UF" (Appendix C.7) with homology specific to the endogenous genomic DNA greater than 1kb upstream of the stop codon (upstream of the 5'

homology sequence). In coordination with the UF primer was the cassette-specific H4neoR, and together these primers straddled the gene-cassette boundary to report on the integration status of the cassette within the *INTS6* locus. The UF/H4neoR primer pair would not produce a signal for aberrant insertion of the knockout cassette since wildtype cells would not have an H4 promoter sequence directly upstream of the target gene, and an atypically inserted cassette would not have the target gene's 5' UTR directly upstream required for fragment amplification. Therefore, a fragment just over 1kb would be produced only if the *IntS6* knockout cassette had successfully recombined with and replaced *INTS6*, otherwise there would be no PCR product (Figure 25, B and C).

Two colonies potentially knocked out for *INTS6* showed a band a little larger than 1kb using the Kpn1F/H4neoR primer pair to signify proper transformation and presence of the knockout cassette, while the UF/H4neoR primer pair identified bands of approximately 1kb in length for all tested colonies indicating successful insertion of the knockout cassette into the *INTS6* locus (Figure 26). Wildtype strains in both cases were void of signal as expected.

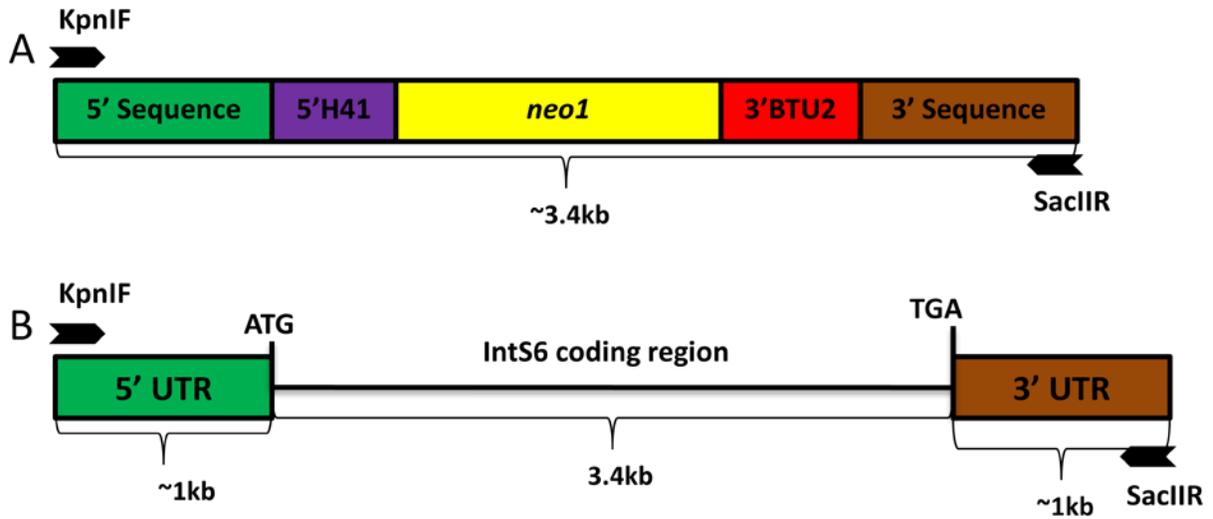


Figure 23. Schematic of expected colony PCR product sizes for testing presence of *INTS6*KO cassette

Possible PCR product sizes used to assess the knockout cassette insertion between different transformation possibilities. A) amplification of the cassette by primers KpnIF and SacIIR to produce a ~3.4kb product irrespective of where it is located in the cell. B) amplification of endogenous *INTS6* to produce a ~5.5kb signal.

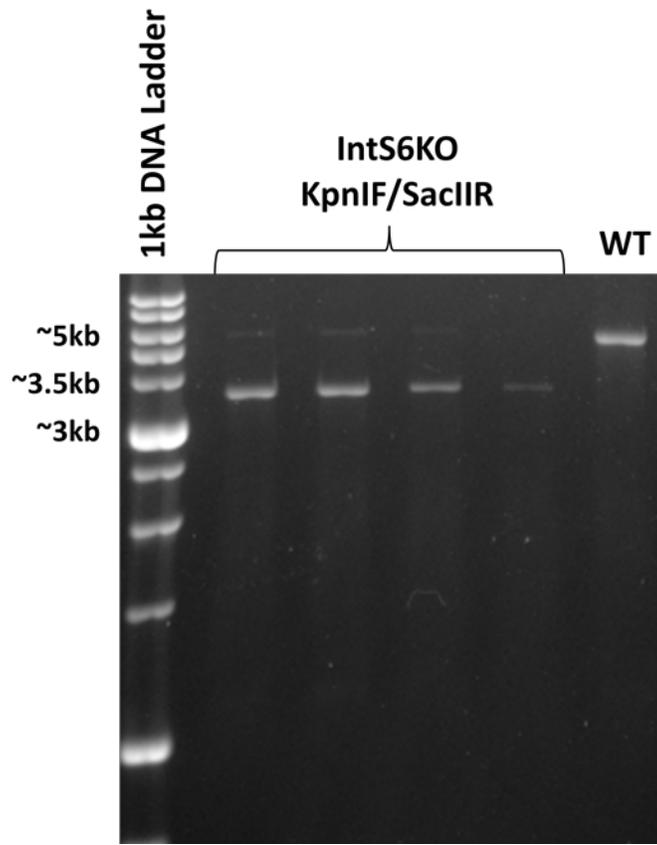


Figure 24. *INTS6*KO colony PCR with KpnIF/SacIIR primer pair

Products of colony PCR from four potential *INTS6* knockout strains and wildtype cells using KpnIF and SacIIR primers were electrophoresed on a 1% agarose gel. Signal around 5kb was present in all lanes (wildtype *INTS6*), while signal near 3.5 kb was only present in transformed strains (*INTS6* knockout cassette).

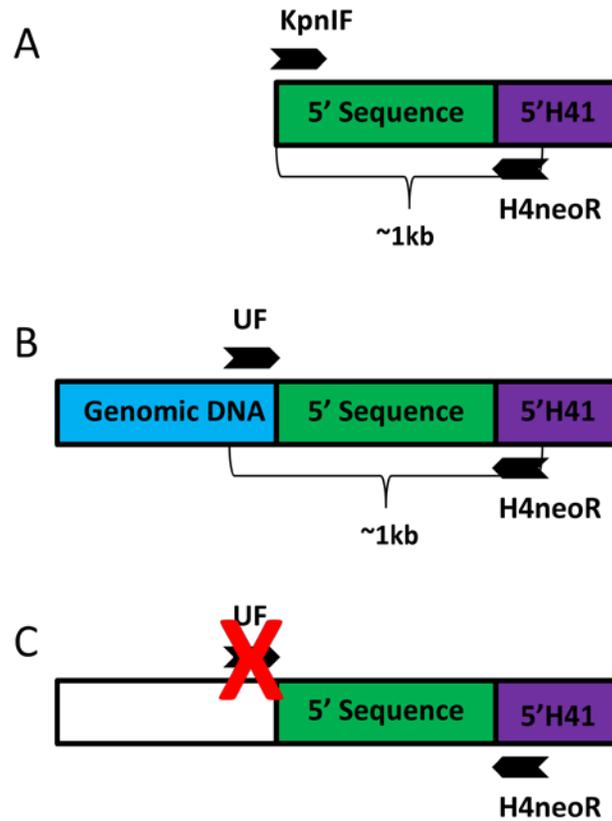


Figure 25. Schematic of expected colony PCR primer binding for testing correct *INTS6*KO integration

Following PCR in the above scenarios, a band around ~1kb representing the 5' homology sequence may be generated. A) Signal will be produced with KpnIF/H4neoR in any cells containing the knockout cassette. B) Signal will only be produced with UF/H4neoR in cells with proper recombination of the knockout cassette with the *INTS6* locus. C) Aberrant insertion of the *IntS6*KO cassette will not produce signal.

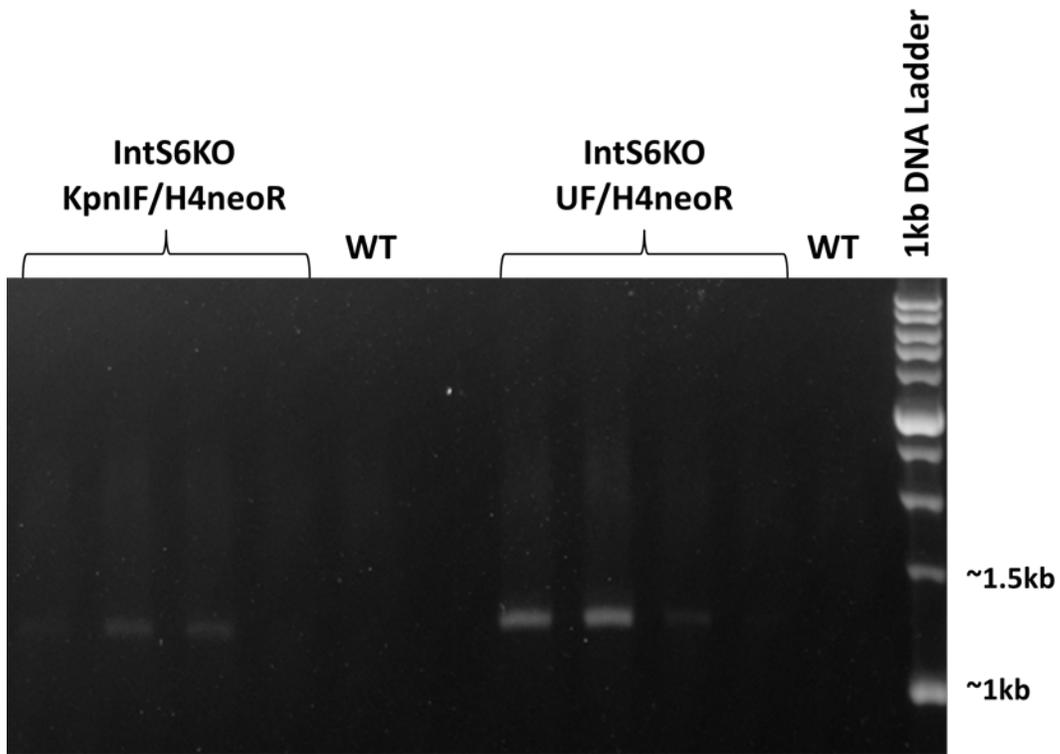


Figure 26. *INTS6*KO colony PCR with KpnIF/H4neoR and UF/H4neoR primer pairs

1% agarose gel loaded with potential *INTS6* knockout colony PCR products. The first set of products was amplified using primers KpnIF and H4neoR to determine only the presence of the *IntS6*KO knockout cassette. The second set was amplified with primers UF and H4neoR to determine the correct insertion of the cassette into the *INTS6* locus.

## Chapter 4: Discussion

In an effort to epitope tag or knock out putative Mediator and Integrator subunits in *T. thermophila*, 1kb sequences homologous to the genomic loci had to be cloned into the respective vectors. The primers used to amplify the 5' and 3' homology regions of the endogenous target genes were designed using raw, un-annotated sequence data based on gene model predictions generated by The Institute for Genomic Research (TIGR) (Stover *et al.*, 2005) retrieved from the *Tetrahymena* Genome Database ([www.ciliate.org](http://www.ciliate.org)). The use of such raw sequence data may negatively implicate downstream applications due to the targeting of incorrect gene loci. Granted this, considerations in primer design were given to ensure primers did not fall within introns and would not amplify a gene product with an internal restriction site corresponding to *KpnI*, *XhoI*, *NotI*, or *SacI* as these are the restriction endonucleases utilized in cassette assembly and would cleave the vector prematurely and incorrectly.

### 4.1 Rpb1 of *Tetrahymena* RNA polymerase II lacks a canonical CTD

Although the canonical heptapeptide sequence of the CTD of Rpb1 is  $Y^1S^2P^3T^4S^5P^6S^7$ , not all repeats conform to this strict consensus sequence. In fact, 6 of the 26 yeast repeats (~23%) and most of the human repeats (~60%) deviate by select amino acid substitutions. The nature of the sequence alignment in Figure 6 prioritized the strings of  $Y^1S^2P^3T^4S^5P^6S^7$  tandem repeats based on their high sequence identity, and since individual substitutions in the repeat sequence are not consistent across species (Appendix C.3), the CTDs were not necessarily aligned as a functional unit. For instance, the first yeast heptapeptide repeat ("5\*" in Figure 6) was aligned 29 residues downstream of the first indicated human repeat due to variations early in

the CTD (Appendix C.3). As a result, modeling the alignment after the CTD of a single organism not only made it difficult to identify the uncharacterized CTD of *Tetrahymena* Rpb1, but potentially skewed calculations which were only inclusive of the elected CTD repeat region.

By examining the relative abundance of amino acids in the CTD of Rpb1 in yeast and humans, it was my intent to uncover similarities in serine, threonine, tyrosine, and proline content in *Tetrahymena* despite its lack of canonical heptapeptad repeats. I initially supposed that a comparable level of these amino acids in *Tetrahymena* would indicate phosphorylation comparable to that of yeast and humans, and formally allow the recruitment of similar CTD-interacting proteins to the *Tetrahymena* Rpb1 CTD. This notion is not accurate, however, since even an abundance of these amino acids would not necessarily presuppose the existence of CTD-interacting proteins in *Tetrahymena* given the lack of CTD sequence regularity and the specificity of CTD-interacting proteins. Without sequence regularity it is difficult to visualize a phospho-serine code as it is not simply arbitrary serine phosphorylation that fosters protein recruitment, but the timely phosphorylation and dephosphorylation at specific positions on the CTD (Bartkowiak *et al.*, 2011).

Although there is no apparent pattern in the CTD sequence of *Tetrahymena* Rpb1, there is a bias for serines in the 2<sup>nd</sup>, 5<sup>th</sup>, or 7<sup>th</sup> position (~67%, 26/39). Furthermore, *Tetrahymena* encodes a homolog to an essential eukaryotic CTD-specific phosphatase Fcp1 (Figure 1) (NCBI; Ghosh *et al.*, 2008). Due to the lack of a defined CTD combined with natural sequence divergence, further examination by computerized analysis may be able to calculate the optimal linker/CTD boundary of *Tetrahymena* RNAPII. Two other options to help elucidate the function of the *Tetrahymena* Rpb1 CTD and the dependence on it by *Tetrahymena* Mediator and Integrator complexes would be to epitope tag it and/or delete it.

In the first method, *Tetrahymena* Rpb1 could be epitope tagged at the N-terminus, perhaps with a glutathione S-transferase (GST) tag, so as not to interfere with the CTD. Affinity purification with immobilized glutathione, the natural substrate for GST, could be used in combination with mass spectrometry to determine if known *Tetrahymena* kinases co-purify (Harper and Speicher, 2011). In addition, chromatin immunoprecipitation could be performed with monoclonal antibodies H5, H14, and 8WG16, which recognize phosphorylated Ser<sup>2</sup> and Ser<sup>5</sup>, and unphosphorylated CTD repeats, respectively (Komarnitsky *et al.*, 2000) to establish whether there is an enrichment of Ser<sup>2</sup>P or Ser<sup>5</sup>P within the CTD of *Tetrahymena* Rpb1 as RNAPII progresses through transcription.

Secondly, the described pBKS-FZZ vector could be engineered to replace the *Tetrahymena* Rpb1 CTD with an FZZ epitope tag. This would enable affinity purification of *Tetrahymena* Rpb1 lacking a CTD. Ideally, this would be used in conjunction with mass spectrometry along with affinity purified wildtype *Tetrahymena* Rpb1 to determine if certain proteins, namely those belonging to the *Tetrahymena* Mediator and Integrator complexes, depend specifically on the CTD for recruitment to Rpb1.

#### **4.2 Med22 AP-MS**

Tandem mass spectrometry of affinity purified Med22 recovered 3 total peptides of the Med22 bait, but no associated peptides. This could have been due to low initial sample size as suggested by the minimal peptide recovery. Med22 would have been inherently enriched through affinity purification, but any loss of protein-protein interaction would have magnified the lack of associated peptides. There are a few speculations that could account for this result, the most elementary of which being that the epitope tag interfered with interactions between Med22 and

other Mediator subunits. This is possible on two accounts: firstly, the tag could have inhibited correct protein folding, and without proper conformation Med22 would have been unable to bind its regular repertoire of proteins, and secondly, the tag itself could have been physically obstructing Med22 from interacting with the rest of the Mediator complex. Alternatively, the low spectral count suggests that the epitope tag may not be accessible to binding by IgG-Sepharose, so perhaps another tagging strategy could be employed.

### 4.3 Med31 AP-MS

Tandem mass spectrometry was performed with affinity purified Med31 to corroborate previous Med31 AP-MS/MS analysis (Fillingham, unpublished). In addition to the formerly acknowledged predicted *Tetrahymena* Mediator orthologs co-purifying with Med31 (Med 4, 7, 11, 17, 20, 21, and 22), this study identified Med3 for the first time as a member of the *Tetrahymena* Med31 interactome to raise the number of purified *Tetrahymena* Mediator subunits to 9 out of the 14 predicted by Bourbon (2008). Med3 is also the first subunit identified in *Tetrahymena* belonging to the tail module of Mediator (Figure 6). This could be a result of a greater challenge in identifying Mediator tail module subunits by homology due to their naturally divergent nature arising from their tendency to interact with species specific DNA-binding factors.

Considering Bourbon predicted an ancient core Mediator "proto-complex" consisting of 17 subunits and has only suggested 13 of such in *T. thermophila*, it is conceivable that a number of proteins identified in the current Med31 AP-MS are *Tetrahymena* Mediator subunits unrecognized due to evolution with minimal selective constraint (Stump and Ostrozhynska, 2013). To further this, the number of significant proteins identified in this Med31 AP-MS (25) is

perhaps coincidentally the same number of subunits recognized in the yeast Mediator complex. Also, Med3 is the most distally located subunit of the tail module (Figure 6), perhaps presupposing the existence of the intermittent Mediator subunits Med1, 2, 5, 9, 14, 15, and 16.

*Tetrahymena* Med31 shares most of the Med31/Med7 interface contact residues as described in yeast by Koschubs *et al.* (2009) (Figure 7). This reinforces the Med31-FZZ AP-MS results identifying predicted *Tetrahymena* Mediator subunit Med7 (Table 4). Despite this lack of sequence conservation for other *Tetrahymena* Mediator subunits, there is another way to potentially identify this co-purifying set of peptides as the bona fide Mediator complex in *Tetrahymena*. Recent x-ray crystallography work in *S. cerevisiae* identified the unique structure of the 4.3Å head module of the yeast Mediator containing three domains: the neck, the moveable jaw, and the unmovable jaw (Imasaki *et al.*, 2011). Electron microscopy previously identified the discrete modules of Mediator, and at a resolution of 4.3Å the 3D structure of the *Tetrahymena* head module could be examined with scrutiny by using scanning electron microscopy and assessed for similar structural qualities (Asturias *et al.*, 1999). If the head module is conserved by its constraints to interact with RNAPII, perhaps a *Tetrahymena* Mediator complex could be identified on the same merits.

#### **4.4 Int11 AP-MS**

Affinity purification-tandem mass spectrometry of Int11-FZZ identified no significant peptides. Probing the Western blots with monoclonal  $\alpha$ -FLAG antibody revealed appropriately sized signal for Int11-FZZ following purification, although it was not definitive. This could have been another situation where the epitope tag interfered with protein-protein interaction, but provided that no bait Int11-FZZ peptides were identified, it is possible that the tag was internalized by the complex and thus became inaccessible to IgG. This would have appeared the

same on a Western blot since protein complexes were disrupted and denatured during subsequent SDS-PAGE, which would have enabled access and probing by the  $\alpha$ -FLAG antibody.

An alternate speculation is that there was an insufficient amount of protein in the sample. To establish this, SDS-PAGE could be combined with protein staining such as silver staining or Coomassie brilliant blue to visualize all proteins in the sample. Considering the potential for low protein concentrations, performing a colloidal Coomassie stain using G-250 instead of R-250 dye would allow for higher sensitivity (Dybala and Metzger, 2012). Another scenario which Coomassie protein staining could contribute toward would be if adequate amount of protein was present but not all protein was tagged, although the 18kDa difference between FZZ-tagged and untagged protein could prove troublesome without adequate protein separation and a suitable protein marker provided the non-specific binding nature of Coomassie.

Transformed *Tetrahymena* cultures had been serially passaged once a week on a typical paromomycin concentration regimen (i.e. 100 $\mu$ g/ml 200 $\mu$ g/ml, 400 $\mu$ g/ml) up to their viable limit of 1200 $\mu$ g/ml. At this stage it was expected that surviving cells had phenotypically assorted with all copies of *INT11* fused to an FZZ epitope tag at the C-terminus and a *neo2* gene. It is improbable that cells would have been able to survive at 1200 $\mu$ g/ml paromomycin without all of their *INT11* loci replaced with the tagging cassette, unless the cells had rejected the FZZ tag along with simultaneous retention of the *neo2* gene. If the FZZ tag induced a dominant-negative phenotype when fused to *INTS6*, the cells would become faced with contradicting selective pressure. Retention of the *neo2* gene would favour antibiotic resistance and cell viability while liberation from the tag would favour phenotypic rescue. In such cases, the affliction could have lead to the selection of a rare event whereby the FZZ tag is lost through homologous recombination with simultaneous retention of the *neo2* gene. Cells would have been visually

indistinguishable and capable of surviving higher concentrations of paromomycin over time as they phenotypically assorted, however, the *neo2* gene would be the only assorting factor, and *INT11* would be left untagged.

#### **4.5 Speculative role of *Tetrahymena* Med31 in meiotic transcription**

Sequence similarity of Mediator subunit Med31 between *Tetrahymena* and other organisms is surprisingly high over most of its primary amino acid sequence, especially considering it is the only Mediator subunit with any sequence similarity in *T. thermophila*. Its N-terminal glutamine (Q)-rich domain suggests *Tetrahymena* Med31 retains a role in transcription activation (Xiao and Jeang, 1998; Escher *et al.*, 2000), but surprisingly, Med31 is non-essential for viability in either *S. cerevisiae* (Fan and Klein, 1994) or *T. thermophila* (Garg and Fillingham, unpublished). Its high sequence conservation in conjunction with its non-essential nature suggests there is selective pressure to retain *MED31*, and insinuates that it may have a redundant role in a separate but fundamental pathway, perhaps interacting with other well-conserved proteins.

Immunofluorescence in *Tetrahymena* revealed that Med31 localizes to micronuclei during conjugation (Garg and Fillingham, unpublished), and while deletion of *MED31* does not affect nuclear division during vegetative growth, mating between two *Tetrahymena* *MED31* knockout strains was accompanied by the inability for cells to complete conjugation (Garg, unpublished). A role for Med31 in development is supported by the gene expression profile of *MED31* (Appendix C.10), and from the timing of expression I postulate a function in meiotic transcription.

Transcription elongation factor IIS (TFIIS) is encoded by *DST1* in yeast (<http://www.yeastgenome.org/>). TFIIS is known to interact with RNA polymerases I, II, and III, and is required for recruiting RNAPII to certain promoters (Schnapp *et al.*, 1996; Wind and Reines, 2000; Guglielmi *et al.*, 2007; Ghavi-helm *et al.*, 2008). BLASTP of yeast TFIIS against the *Tetrahymena* Genome Database ([www.ciliate.org](http://www.ciliate.org)) contains an annotation for TFIIS (TTHERM\_00691200).

In yeast, TFIIS acts in conjunction with Med31 for cell viability, as the deletion of *DST1* is synthetic lethal with *med31Δ*. Interestingly, TFIIS has an Rpb1-binding domain specifically required for viability only in the absence of Med31, and is otherwise dispensable under wildtype Med31 conditions (Guglielmi *et al.*, 2007). In essence it seems that both TFIIS and Med31 are able to recruit RNAPII, perhaps in a redundant manner by interacting with the same part of RNAPII. In *dst1Δ/med31Δ* yeast strains, supplementation with the Dst1 Rpb1-binding domain alone was sufficient to rescue the synthetic lethal phenotype. Conversely, when the ability of TFIIS to recruit RNAPII directly was eliminated by mutation of its Rpb1-binding domain in wildtype Med31 strains, it appeared that TFIIS was still able to do so via recruitment of Med31.

Although current *Tetrahymena* purification data does not reveal an interaction between Med31 and TFIIIs, if Med31 can be found to interact with this established active participant in meiosis, the micronuclear localization of Med31 during conjugation and the concomitant disruption of conjugation imparted by a *MED31* knockout would conceivably support a role for Med31 in meiotic transcription. To draw a parallel with the aforementioned yeast study, both Med31 and TFIIS could be knocked out in *Tetrahymena* to determine whether this is also a lethal combination.

#### 4.6 Tagging *INTS6* with a C-terminal FZZ epitope tag in *T. thermophila*

A full FZZ-tagging vector specific to *Tetrahymena INTS6* was constructed and verified through sequencing, however, multiple attempts at tagging this gene through biolistics were unsuccessful, and cells would perish during transfer to paromomycin concentrations of 100µg/ml. I speculate that cell death was caused by the prevention of recruitment of Integrator subunit IntS6 to the Integrator complex imparted by the epitope tag, either through interference of protein-protein interactions or conformational change to IntS6, compounded with the essential requirement for Integrator subunits for cell viability. Therefore, it may be possible to overcome the encumbrance of the FZZ tag on IntS6 if it was added to the N-terminus of *INTS6* without disrupting upstream regulatory regions.

IntS6 was previously recognized as the candidate tumor suppressor "deleted in cancer 1" (DICE1) in human non-small cell lung carcinoma (Wieland *et al.*, 1999). It is interesting that an Integrator subunit, a deletion of which would inhibit proper snRNA formation, could be implicated in tumorigenesis. Absence of a signal motif suggests intracellular functionality (Chen and Wagner, 2010), but the presence of an N-terminal von Willebrand factor A is a regular feature of proteins with diverse roles, and so suggests a similar role for *Tetrahymena* IntS6 (Whittaker and Hynes, 2002). Interrupted oogenesis and increased apoptosis in *C. elegans* strains knocked out for the *INTS6* homolog *DIC-1* (Han *et al.*, 2006) compounded by the vital necessity for *INTS6* in *T. thermophila* (current study) promotes the hypothesis that IntS6 is essential for cellular development, and that any factor inhibiting recruitment of IntS6 to the Integrator complex could lead to cell inviability.

#### 4.7 Future prospective of Integrator knockout

The gene expression profile for *INTS6* in *Tetrahymena* (Appendix C.10) peaks in very early-mid conjugation, supporting a role in 3' snRNA processing, however future experiments are required to confirm this potential role. Furthermore, the temporal occurrence of this peak corresponds with a peak observed in the gene expression profile of *Tetrahymena DSS1* (Appendix C.10), which reinforces the importance of the *T. thermophila* Dss1 purification data and in turn suggests that a *Tetrahymena* Integrator complex may exist.

Following colony PCR with the KpnIF and SacIIR primer pair (Figure 24), the appearance of a larger ~5.4kb band indicative of wildtype *INTS6* in the same strains where correct cassette integration had occurred (Figure 26) supports an essential role for IntS6 as it suggests that despite being under such selective pressure that cells are dying, not all macronuclear copies of *INTS6* were able to be replaced with the *neoI* gene. However, as previously mentioned, this faint higher molecular weight band could represent micronuclear *INTS6*. Also, these cells did not flourish in paromomycin concentrations above 200µg/ml, and would survive only up to 400µg/ml. This is a very low drug concentration as homozygous cells with a *neo* gene on all of its 45 copies of a gene survived up to and beyond 1mg/ml paromomycin (current study). This suggests that IntS6 is essential for cell viability, and as increasing drug concentration increased the selective pressure on the cells to replace more copies of *INTS6* with *neoI*, a minimum threshold for *INTS6* was attained and cells perished. These findings are supported by *INTS6/DIC-1* knockout experiments in *C. elegans* (Han *et al.*, 2006).

To fully elucidate the functional consequence of knocking out the *Tetrahymena* Integrator complex, it is necessary to evaluate the extent of *INTS6* knockout in the macronucleus to determine whether it is essential for cell viability. The immediate undertaking would be to test

whether the knockout of *INTS6* is a complete null by running a Southern blot using probes against *Tetrahymena INTS6* to detect its presence. Considering there are approximately 45 copies of *INTS6* in the wildtype *Tetrahymena* macronucleus and phenotypic assortment did not advance considerably, it is likely that native *INTS6* was still present within the cell and would be detectable in a Southern blot. If micronuclear *INTS6* copies are interfering with Southern blot analysis, the nuclei could be centrifugally separated based on density. Alternatively, reverse transcriptase (RT)-PCR on isolated *Tetrahymena INTS6KO* RNA using primers internal to *INTS6* could be employed to determine gene expression. These primers could even be engineered to flank the intron in *INTS6*, which would allow distinction of PCR template between micronuclear DNA and cDNA synthesized from intron-free macronuclear mRNA for further distinction.

To determine whether snRNA processing was affected in *T. thermophila* provided the amount of *INTS6* present determined by a Southern blot, a Northern blot using probes against *Tetrahymena* U1 and U2 snRNA could be used to detect whether snRNA expression is affected in an *INTS6* knockout.

## **Conclusion**

The goal of this study was to elucidate and characterize RNA polymerase II-interacting complexes Mediator and Integrator in the ciliated protozoan *Tetrahymena thermophila* for the first time by performing affinity purification coupled with mass spectrometry. RNA polymerase II is the core protein complex essential for production of messenger RNA in eukaryotes. In most eukaryotes, the largest subunit of RNAPII, Rpb1, contains a canonical C-terminal domain that acts as a platform for protein-protein interactions, and has been implicated in a multitude of

mRNA processing events. Sequence alignment of Rpb1 of RNAPII was unable to identify the canonical Y<sup>1</sup>S<sup>2</sup>P<sup>3</sup>T<sup>4</sup>S<sup>5</sup>P<sup>6</sup>S<sup>7</sup> heptapeptide repeat structure for the CTD in *Tetrahymena thermophila*. Two RNAPII-interacting protein complexes, Mediator and Integrator, elicit their influence on RNAPII during transcription through direct binding to the CTD of Rpb1. Recent affinity purifications coupled with mass spectrometry in *T. thermophila* have uncovered potential Mediator and Integrator subunits in this organism lacking a canonical CTD on RNAPII.

Mediator subunit Med31, was identified through BLASTP as the only Mediator subunit with any sequence conservation in *T. thermophila* despite being the only identified Mediator subunit that is not essential for viability in either *T. thermophila* or *S. cerevisiae*. Despite a primary sequence identity of 33% between yeast and *Tetrahymena* Med31, *Tetrahymena* Med31 was not able to rescue the slow growth phenotype resulting from a knockout of *MED31* in yeast, nor did it capitulate an interaction with yeast subunit Med8. Affinity purification coupled with mass spectrometry in *T. thermophila* identified Med3 for the first time as an interacting partner of Med31 and a potential *Tetrahymena* Mediator subunit, while AP-MS using other Mediator or Integrator subunits was inconclusive. Exact gene replacement of *INTS6* loci with a *neo1* knockout cassette was confirmed through colony PCR, however, the inability of these cells to thrive in paromomycin concentrations above 200µg/ml supports an essential role for the *INTS6* gene product in *T. thermophila*. It was my hope that the data described can be extrapolated in the long-term to determine whether these complexes serve a purpose in *Tetrahymena* comparable to higher eukaryotes to gain insight into the functional consequence of life without a canonical CTD. The CTD of *Tetrahymena* Rpb1 therefore requires further investigation into potential phosphorylation and protein recruitment.

## Appendix A: Fms1 and Vps75

To maintain the organization, stability, and integrity of DNA while meeting the size limitations imposed by the nucleus, eukaryotic DNA is strategically coiled into a higher order structure called chromatin (Shahbazian and Grunstein, 2007). In order to achieve this arrangement, an octomer of histone proteins (two H2A-H2B dimers and an H3-H4 tetramer), is wrapped with 147 base pairs of DNA to form a nucleosome (Luger *et al.*, 1997). Covalent histone modification such as acetylation by histone acetyltransferases (HATs) plays a major role in chromatin remodeling necessary for effective transcription, replication, and repair by relaxing the tightly wound DNA and granting accessibility by the appropriate factors (Shahbazian and Grunstein, 2007).

In yeast, one such histone acetyltransferase is Rtt109, which acetylates histone H3 on lysines 9, 27, and 56 (H3K9, H3K27, and H3K56, respectively). The histone chaperone protein Vps75 (Selth and Svejstrup, 2007) forms a stable complex with Rtt109 to stabilize it and enhance its histone acetyltransferase activity (Krogan *et al.*, 2006; Fillingham *et al.*, 2008). Raw affinity purification-mass spectrometry data from *S. cerevisiae* reveals that Vps75 co-purifies with polyamine oxidase Fms1 (tap.med.utoronto.ca). Fms1 oxidizes polyamine spermine to generate spermidine, which is necessary for the hypusine modification of eukaryotic translation initiation factor 5A (eIF-5A); the only protein known to carry this unique amino acid (Landry and Sternglanz, 2003; Zanelli *et al.*, 2006; Wolff *et al.*, 2007; Park *et al.*, 2010). Polyamines such as spermine can incorporate covalently into chromatin and repress transcription (Tabor and Tabor, 1984; Igarashi and Kashiwagi, 2000), and histone acetylation has been shown to overcome this

repression (Pollard *et al.*, 1999). Perhaps Fms1 interacts with Vps75 to utilize the polyamines as they are freed from chromatin following histone acetylation by Rtt109.

Interestingly, a potential Rtt109 consensus sequence consisting of lysine-serine-threonine (KST) common within the most highly acetylated lysine residues of H3 also exists within Fms1, suggesting that perhaps Fms1 can be acetylated by Rtt109 (Fillingham, unpublished). It was therefore of interest to investigate whether Fms1 interacts with the Rtt109-Vps75 complex through Vps75, or whether it was acetylated by Rtt109 on the KST consensus sequence, effectively linking biosynthesis with transcriptional regulation.

To recapitulate the Fms1-Vps75 interaction, a C-terminal 13xmyc epitope tag was fused to Fms1 in a Vps75-TAP background. PCR amplification of plasmid pFA6a-13myc-kanMX6 as described by Longtine *et al.* (1998) using primers with 40bp homologous to the 3' end of *FMS1* flanking the stop codon and 20bp homologous to the vector synthesized a tagging cassette that was able to insert itself into the haploid yeast genome through homologous recombination to generate a homozygous Fms1-13myc strain. The kanMX6 module carried on the tagging vector conferred resistance to G418/geneticin, and was of *E. coli* in origin as to prevent aberrant homologous recombination into the yeast genome (Longtine *et al.*, 1998). The tagging vector was introduced into wildtype (OB+) and Vps75-TAP strains and a one-step affinity purification using IgG-sepharose was performed followed by a Western blot probed with  $\alpha$ -FLAG (1:2000) and  $\alpha$ -myc (1:3000).

Unfortunately, while the Western blot suggested Vps75 and Fms1 were tagged appropriately, as determined by the input fragment sizes (Vps75-TAP=51kDa and Fms1-13myc=73.4kDa), and that sufficient Vps75 was affinity purified, there was not a signal specific to the appropriate strains corresponding to Fms1-13myc in the affinity purification samples

(Figure A1). A signal for  $\alpha$ -myc appeared in all affinity purified samples despite the fact that  $OB^+$  and Vps75-TAP strains are free of a myc tag as supported by the input. Until the Fms1-Vps75 interaction can be recapitulated, this project is suspended.

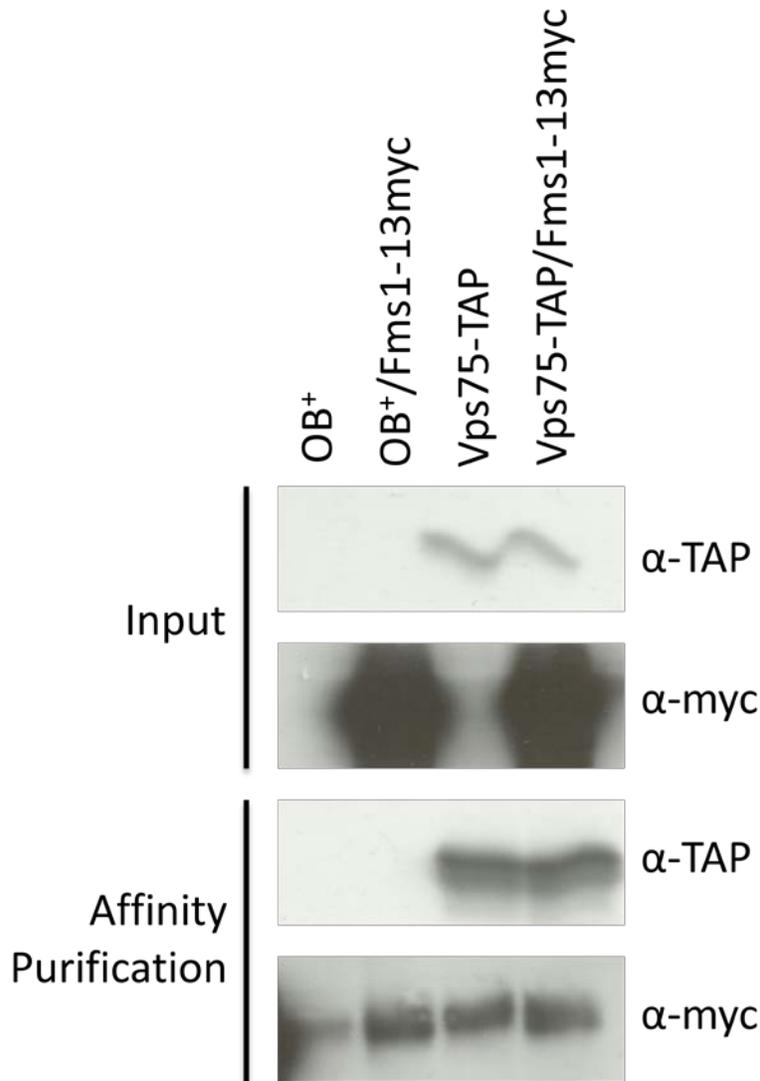


Figure A1. Western blot of Vps75-TAP affinity purification. Western blot depicting input and affinity purified  $OB^+$ ,  $OB^+$ /Fms1-13myc, Vps75-TAP, and Vps75-TAP/Fms1-13myc. Non-specific binding was present in all lanes of affinity purified sample probed with  $\alpha$ -myc. Blots were probed with  $\alpha$ -FLAG (dilution 1:1000, exposure=1 second) and  $\alpha$ -myc (dilution 1:3000, exposure=5 minutes).

If the interaction can be confirmed, an *in vitro* HAT assay would be performed. The ~1,500bp coding sequence of *FMS1* would be PCR amplified with primers pertaining to pET28a expression vector so that product *FMS1* could be cloned in. This would add a 6x histidine (6x HIS) epitope tag to facilitate the purification of Fms1 from *E. coli* for the detection of Fms1 acetylation. The importance of Fms1 acetylation in spermidine biosynthesis could be elucidated by subsequent site directed mutagenesis of the KST consensus sequence followed by a reevaluation of histone acetyltransferase activity on Fms1.

## Appendix B: Rtt109

Histone chaperones Asf1 and Vps75 form a poorly understood complex with fungal histone acetyltransferase (HAT) Rtt109 in the process of histone H3 acetylation at positions 9 and 56 (H3K9ac and H3K56ac, respectively) during chromatin remodeling for genetic stability and transcriptional regulation (Bannister and Kouzarides, 2011). Investigation into the role of the evolutionarily conserved C-terminal tail of Rtt109 (Rtt109C) and of its auto-acetylation at lysine 290 (K290) generated clearer insight into Asf1 and Vps75 regulation of Rtt109-mediated acetylation.

Eukaryotic  
Cell

### The Carboxyl Terminus of Rtt109 Functions in Chaperone Control of Histone Acetylation

Ernest Radovani, Matthew Cadorin, Tahireh Shams, Suzan El-Rass, Abdel R. Karsou, Hyun-Soo Kim, Christoph F. Kurat, Michael-Christian Keogh, Jack F. Greenblatt and Jeffrey S. Fillingham  
*Eukaryotic Cell* 2013, 12(5):654. DOI: 10.1128/EC.00291-12.  
Published Ahead of Print 1 March 2013.

My contribution to this paper involved transforming yeast strains deleted for *RTT109* and containing a C-terminally TAP-tagged Vps75 (*rtt109Δ/Vps75-TAP*) with different iterations of Rtt109. Firstly, wildtype *RTT109* contains a lysine at residue 290 and a basic patch at its carboxy terminus. One Rtt109 alternative was K290R, which encoded Rtt109 with an arginine (R) in place of the lysine 290 residue and was unable to be acetylated, but still contained a positive charge characteristic of lysine. Alternately, K290Q had glutamine (Q) substituted for lysine at this position and mimicked a constitutively acetylated lysine. I generated a second set of each of these three versions of Rtt109 using a primer that would truncate Rtt109 to effectively delete the carboxy-terminal basic. Each version of *RTT109* was cloned into the 12myc-pRb415

plasmid for the addition of an N-terminal 12xmyc epitope tag to Rtt109. I then performed affinity purifications and subsequent Western blots to determine if Rtt109 lysine 290 was required for its interaction with Vps75 and how the basic carboxy terminal played into this interaction. This information was also used as a cross reference during downstream HAT assays to compare the ability of Vps75 to bind mutated K290 with H3K9 and H3K56 acetylation status to determine how the K290 position was implicated in acetylation.

My affinity purifications helped reveal that while mutating Rtt109 lysine 290 to arginine or glutamine did not affect binding of Vps75 to full-length Rtt109, K290R and K290Q strains showed significantly reduced acetylation of H3K9 (Figure B1).

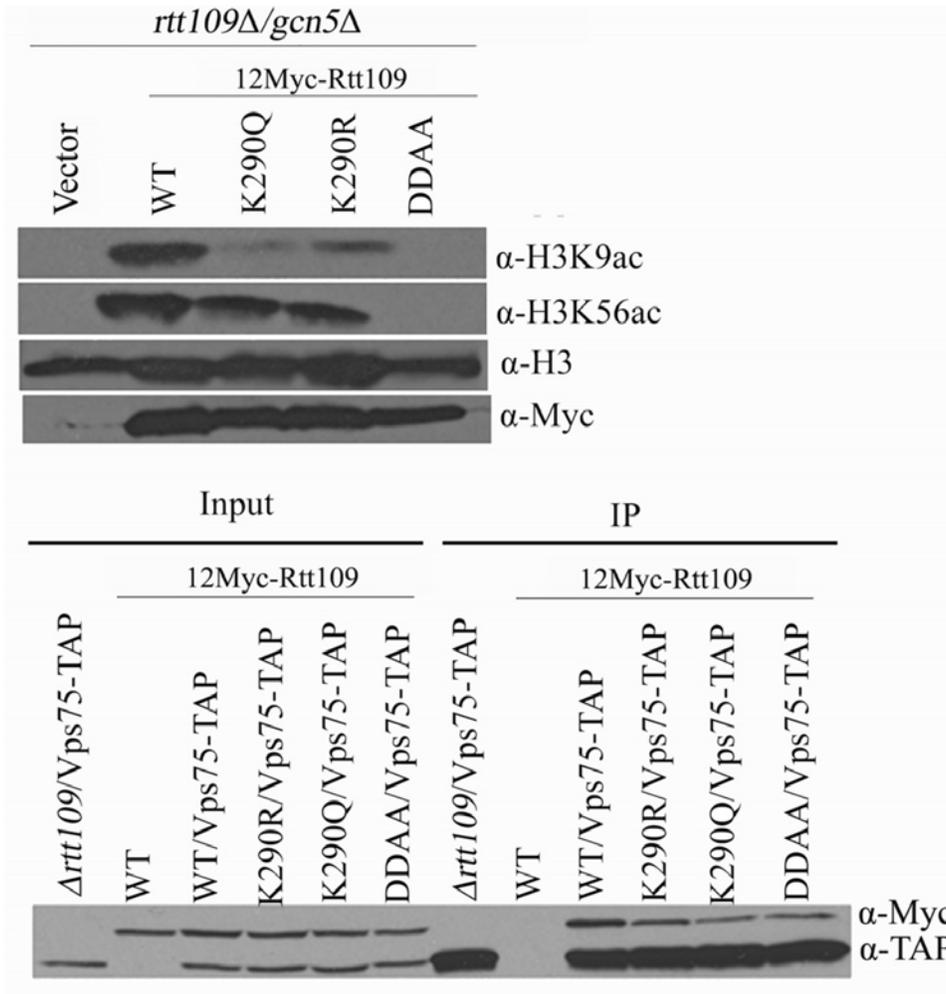


Figure B1. Rtt109 lysine 290 is necessary for optimal H3K9 acetylation. Lysine 290 of Rtt109 is important for *in vivo* H3K9 acetylation by Rtt109, but does not affect binding of Vps75 to Rtt109.

In the other strains bearing a deletion of the basic patch at the carboxy terminus of Rtt109, binding by Vps75 was again not affected by the truncation in conjunction with lysine 290 modification, however, acetylation of H3K56 in K290R and K290Q mutants was reduced, suggesting this basic patch is necessary for enhanced H3K56ac (Figure B2). Since binding by Vps75 was not affected in either case, but K290 played an important role in acetylation status,

this paper showed that *in vivo* Vps75-mediated H3 acetylation by Rtt109 is affected by lysine 290; the first evidence for an *in vivo* role for Vps75 in H3K56 acetylation.

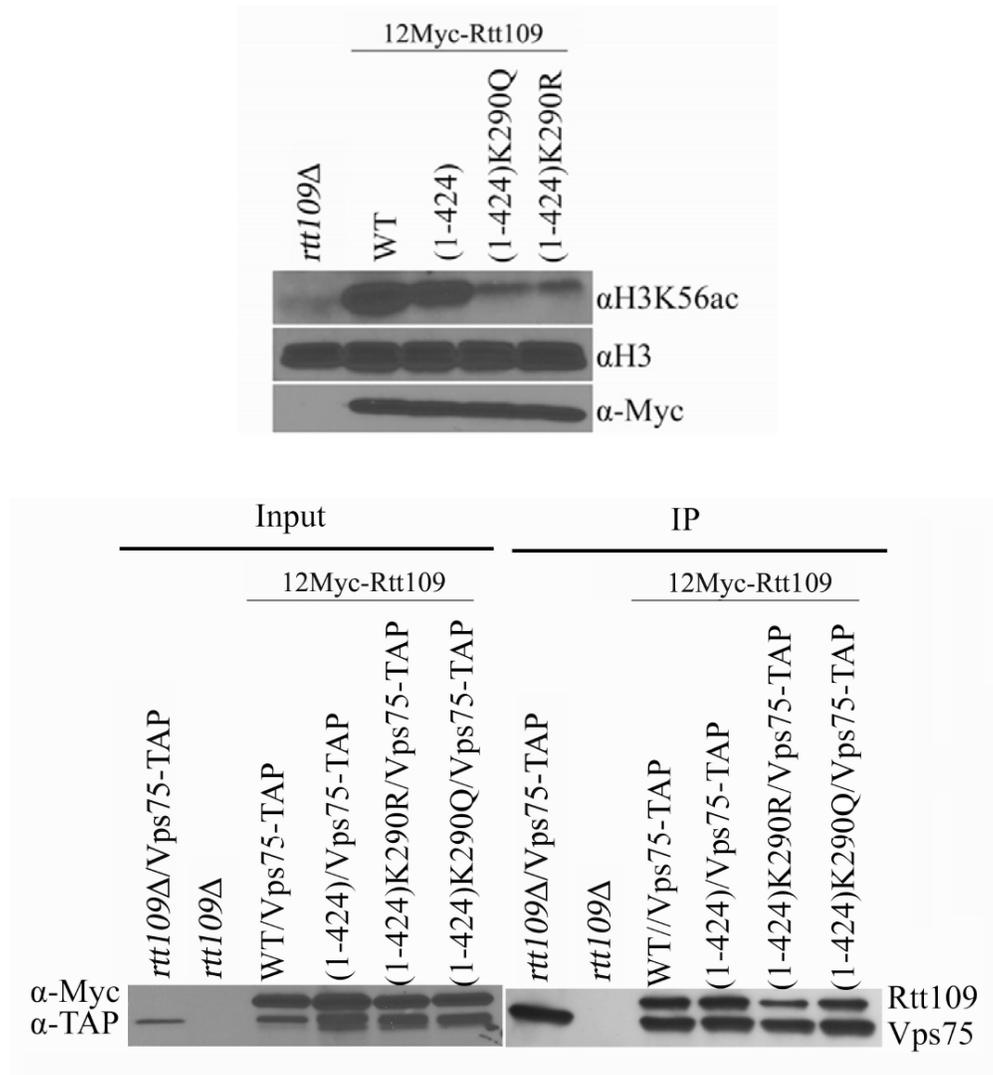


Figure B2. Rtt109 lysine 290 is important for optimal H3K56 acetylation. Lysine 290 of Rtt109 is important for optimal H3K56 acetylation by Rtt109 *in vivo* in the absence of the carboxy terminal patch on Rtt109, but does not affect binding of Vps75 to Rtt109.

The fundamental importance of this study arises from Rtt109 being involved in the pathogenicity of *C. albicans* (Lopes *et al.*, 2010). Its lack of sequence homology to any identified histone acetyltransferase Rtt109 makes it an excellent candidate target for anti-fungal therapeutics, and elucidating its underlying mechanisms could be a step forward in achieving a therapeutic strategy.



## The Carboxyl Terminus of Rtt109 Functions in Chaperone Control of Histone Acetylation

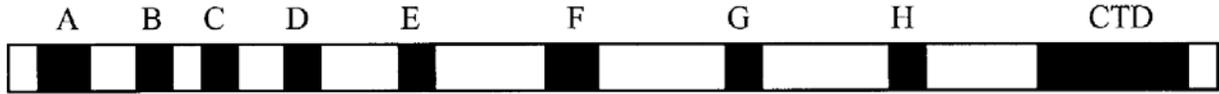
Ernest Radovani,<sup>a</sup> Matthew Cadorin,<sup>a</sup> Tahireh Shams,<sup>a</sup> Suzan El-Rass,<sup>a</sup> Abdel R. Karsou,<sup>a</sup> Hyun-Soo Kim,<sup>d</sup> Christoph F. Kurat,<sup>e</sup> Michael-Christopher Keogh,<sup>d</sup> Jack F. Greenblatt,<sup>b,c,e</sup> Jeffrey S. Fillingham<sup>a</sup>

Department of Chemistry and Biology, Ryerson University, Toronto, Ontario, Canada<sup>a</sup>; Banting and Best Department of Medical Research, Donnelly Centre, University of Toronto, Toronto, Canada<sup>b</sup>; Department of Molecular Genetics, University of Toronto, Toronto, Canada<sup>c</sup>; Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York, USA<sup>d</sup>; The Donnelly Center, University of Toronto, Toronto, Ontario, Canada<sup>e</sup>

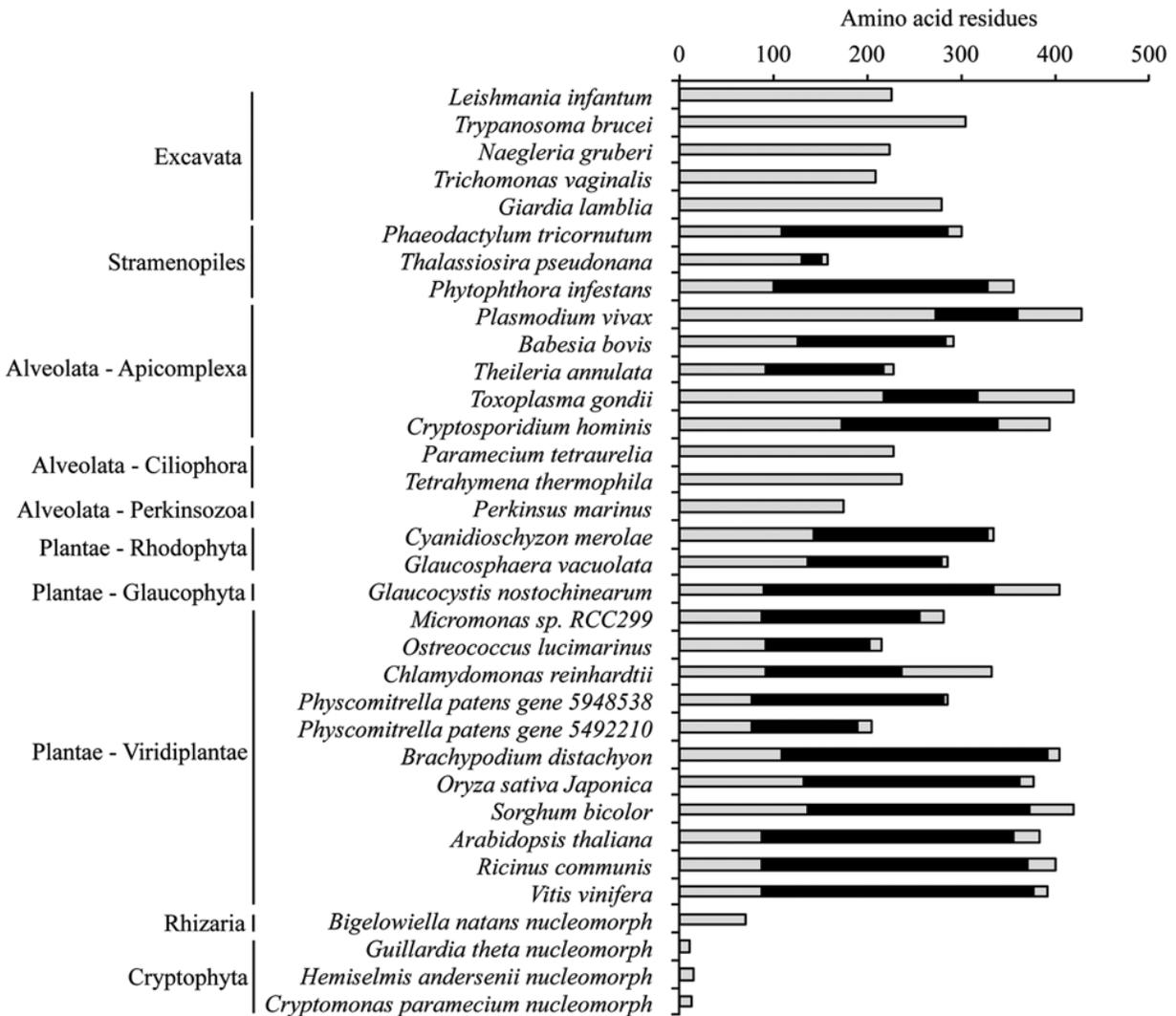
Rtt109 is a fungal histone acetyltransferase (HAT) that catalyzes histone H3 acetylation functionally associated with chromatin assembly. Rtt109-mediated H3 acetylation involves two histone chaperones, Asf1 and Vps75. *In vivo*, Rtt109 requires both chaperones for histone H3 lysine 9 acetylation (H3K9ac) but only Asf1 for full H3K56ac. *In vitro*, Rtt109-Vps75 catalyzes both H3K9ac and H3K56ac, whereas Rtt109-Asf1 catalyzes only H3K56ac. In this study, we extend the *in vitro* chaperone-associated substrate specificity of Rtt109 by showing that it acetylates vertebrate linker histone in the presence of Vps75 but not Asf1. In addition, we demonstrate that in *Saccharomyces cerevisiae* a short basic sequence at the carboxyl terminus of Rtt109 (Rtt109C) is required for H3K9ac *in vivo*. Furthermore, through *in vitro* and *in vivo* studies, we demonstrate that Rtt109C is required for optimal H3K56ac by the HAT in the presence of full-length Asf1. When Rtt109C is absent, Vps75 becomes important for H3K56ac by Rtt109 *in vivo*. In addition, we show that lysine 290 (K290) in Rtt109 is required *in vivo* for Vps75 to enhance the activity of the HAT. This is the first *in vivo* evidence for a role for Vps75 in H3K56ac. Taken together, our results contribute to a better understanding of chaperone control of Rtt109-mediated H3 acetylation.

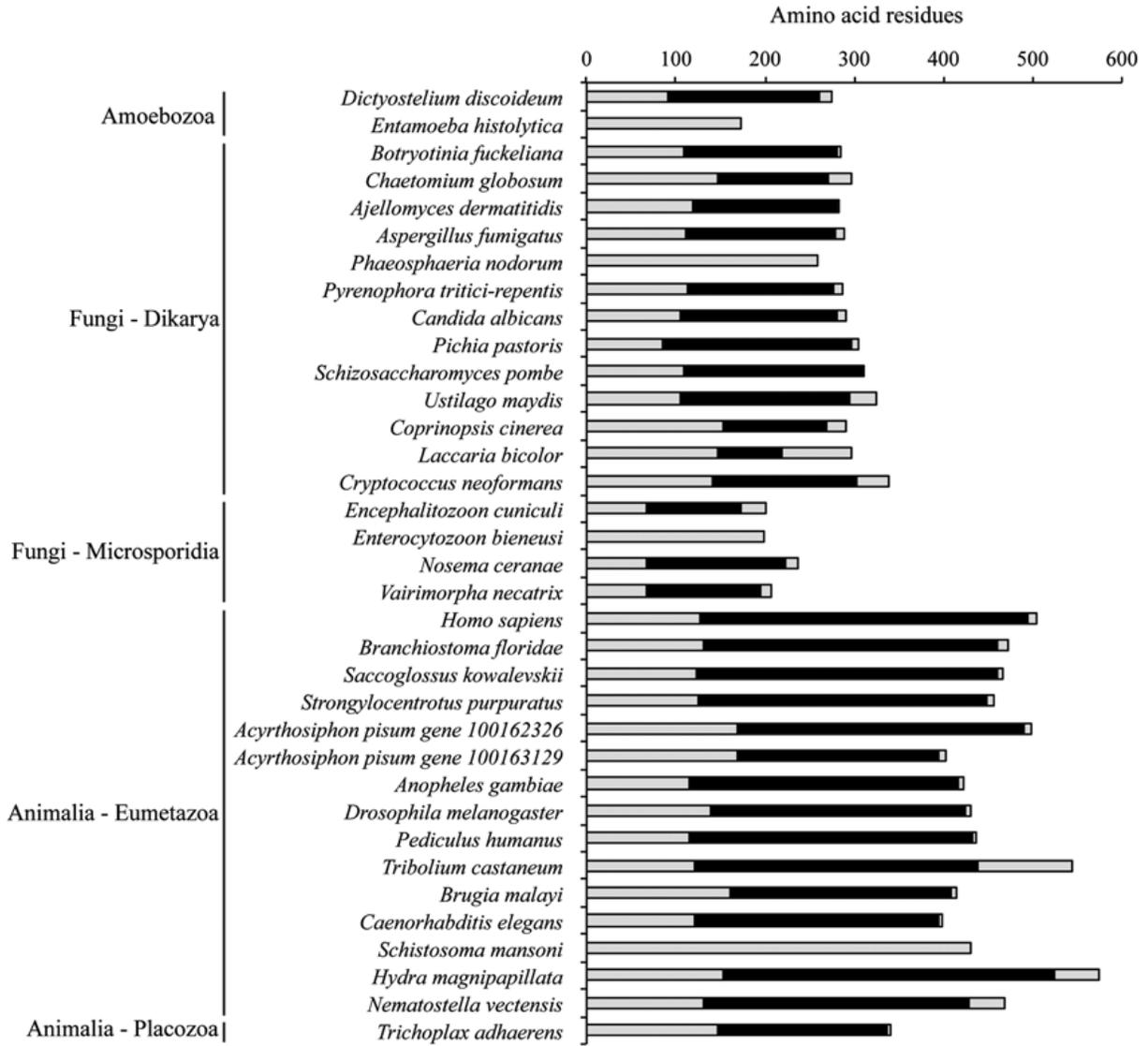
## Appendix C: Supplementary Material

C.1. Schematic of Rpb1 depicting the eight homology regions A-H and the CTD (Matheny *et al.* 2002).



C.2. Length of Rpb1 C-terminal extension following last conserved homology region from a variety of species (Stump and Ostrozhynska, 2013).







#### C.4. TTHERM accession numbers

<b>Gene</b>	<b>TTHERM accession number</b>
Med3	TTHERM_00490630
Med4	TTHERM_00691210
Med17	TTHERM_00780600
Med20	TTHERM_00922930
Med22	TTHERM_00670380
Med31	TTHERM_00355460
Int2	TTHERM_00532780
Int4	TTHERM_00467840
IntS6	TTHERM_01243440
Int9	TTHERM_01159920
Int11	TTHERM_00339790
Rpb1	TTHERM_00538940
Rpn15/Dss1	TTHERM_00227230
Fcp1	TTHERM_00277390
Brca2	TTHERM_00437260

#### C.5. Composition of all media, buffers, and solutions used

<b>Item</b>	<b>Composition</b>
1% Agarose Gel (w/v) (50ml)	0.5g agarose  50ml 1xTBE  5µl ethidium bromide (EtBr)  (10mg/ml)
0.5M Ammonium Hydroxide (NH <sub>4</sub> OH)	1ml 14.5M NH <sub>4</sub> OH  28ml ddH <sub>2</sub> O
10% APS (Ammonium persulfate) (w/v)	0.1g ammonium persulfate  1ml ddH <sub>2</sub> O
AP Lysis Buffer	10ml AP wash buffer  1 tablet complete protease inhibitor  (Roche)

	50µl phenylmethylsulfonyl fluoride (PMSF)
AP Wash Buffer	0.5ml 1M Tris pH 8.0 1.5ml 5M NaCl 0.5ml 10% NP40 47.5ml ddH <sub>2</sub> O
1M CaCl <sub>2</sub> (1L)	Determine which of the following hydrated forms is available: CaCl <sub>2</sub> = 110.99g/mol CaCl <sub>2</sub> · 2 H <sub>2</sub> O (Dihydrate) = 147.02g/mol CaCl <sub>2</sub> · 4 H <sub>2</sub> O (Tetrahydrate) = 183.04g/mol CaCl <sub>2</sub> · 6 H <sub>2</sub> O (Hexahydrate) = 219.08g/mol To 1 mole of CaCl <sub>2</sub> add ddH <sub>2</sub> O to 1L
2mM CaCl <sub>2</sub> /20mM Tris	100µl 1M CaCl <sub>2</sub> 1ml 1M Tris pH 8.0 48.9ml ddH <sub>2</sub> O
0.5M EDTA, Iron (III) Sodium Salt pH 8.0 (500ml)	91.78g Na <sub>2</sub> EDTA (367.1g/mol) ddH <sub>2</sub> O to 500ml, pH to 8.0 with NaOH
Fillingham's 2x Lysis Buffer (50ml)	2ml 1M Tris pH 8.0

	6ml 5M NaCl 50µl 1M MgCl <sub>2</sub> 42ml ddH <sub>2</sub> O
1M MgCl <sub>2</sub> (M.W.=203.3g/mol) (100ml)	20.33g MgCl <sub>2</sub> ddH <sub>2</sub> O to 100ml
1% Milk Solution (50ml)	10ml 5% milk solution 40ml 1x PBS
5% Milk Solution (BLOTTO) (w/v) (100ml)	5g skim milk powder 100ml PBS
Miniprep Solution 1 (400ml)	10ml 1M TRIS pH 8.0 8ml 0.5 M EDTA 9ml 40 % glucose 373ml ddH <sub>2</sub> O
Miniprep Solution 2 (10ml)	1ml 10% SDS 2ml 1N NaOH 7ml ddH <sub>2</sub> O
Miniprep Solution 3 (500ml)	147.2g 3M potassium acetate 120.1g acetic acid 500ml ddH <sub>2</sub> O
100mM NaCl Wash Buffer (IPP100)	500µl 1M Tris pH 8.0 1ml 5M NaCl 500µl 10% NP-40 48ml ddH <sub>2</sub> O

300mM NaCl Wash Buffer (IPP300)	500µl 1M Tris pH 8.0 3ml 5M NaCl 500µl 10% NP-40 46ml ddH <sub>2</sub> O
5M NaCl (500ml)	146.1g NaCl ddH <sub>2</sub> O to 500ml
10% NP-40 (v/v)	2.5ml NP-40 22.5ml ddH <sub>2</sub> O
10x PBS pH 7.3 (1L)	82g NaCl 2.64g NaH <sub>2</sub> PO <sub>4</sub> 16g Na <sub>2</sub> HPO <sub>4</sub> ddH <sub>2</sub> O to 1L, pH 7.3
1x PBST (500ml)	500ml 1x PBS 250µl Tween 20
13% PEG (w/v) (100ml)	13g PEG8000 1M NaCl to 100ml
100mM PMSF (10ml)	0.1742g PMSF 10ml isopropanol
Ponceau (0.1% w/v) (1L)	1g Ponceau S 50ml acetic acid ddH <sub>2</sub> O to 1L
Proteinase K Buffer	100µl 10x ThermoPol buffer (New England BioLabs)

	<p>10<math>\mu</math>l Proteinase K (50<math>\mu</math>g/ml) (New England BioLabs)</p> <p>890<math>\mu</math>l ddH<sub>2</sub>O</p>
2x SDS Laemmli Sample Buffer	<p>3g SDS</p> <p>5ml beta-mercaptoethanol</p> <p>10ml 100% glycerol</p> <p>6ml 2M Tris-HCL pH 6.8</p> <p>50mg bromophenol blue</p> <p>ddH<sub>2</sub>O to 100ml</p>
SPP (1L)	<p>60mg sequestrin (Sigma)</p> <p>2g bacto yeast extract</p> <p>20g proteose peptone</p> <p>4g glucose</p> <p>ddH<sub>2</sub>O to 1L</p>
SPP+PSF (Penicillin/Streptomycin/Fungizone) (1L)	<p>1L SPP</p> <p>500 <math>\mu</math>l PSF (100x)</p>
5% Stacking Gel (5ml)	<p>3.5ml ddH<sub>2</sub>O</p> <p>0.625ml 1M Tris pH 6.8</p> <p>0.95ml acrylamide 29:1</p> <p>0.05ml 10% SDS</p> <p>3.75<math>\mu</math>l TEMED</p> <p>31.25<math>\mu</math>l 10% APS</p>
10x TBE pH 8.0 (4L)	10g Tris base

	<p>5.5 Boric acid</p> <p>0.93g EDTA</p> <p>4L ddH<sub>2</sub>O</p> <p>Titrate with HCl until pH=8.0</p>
<i>Tetrahymena</i> Lysis Solution (500ml)	<p>210g urea</p> <p>35ml 5M NaCl</p> <p>5ml 1M Tris pH 7.4</p> <p>10ml 0.5M EDTA</p> <p>50ml 10% SDS</p> <p>ddH<sub>2</sub>O to 500ml</p>
1x TEV Cleavage Buffer	<p>500μl 1M Tris pH 8.0</p> <p>1ml 5M NaCl</p> <p>500μl 10% NP-40</p> <p>50μl 0.5M EDTA</p> <p>48ml ddH<sub>2</sub>O</p>
1M Tris-HCl pH 8.0 (500ml)	<p>60.55g Tris</p> <p>ddH<sub>2</sub>O to 500ml, pH to 8.0</p>
10mM Tris pH 7.4 (1L)	<p>1.21g Tris</p> <p>ddH<sub>2</sub>O to 1L, pH to 7.4</p>
4x Western Running Buffer (4L)	<p>230.4g glycine</p> <p>48g Tris</p> <p>16g SDS</p> <p>ddH<sub>2</sub>O to 4L</p>

Western Transfer Buffer (2L)	400ml methanol 6.05g Tris 28.84g glycine ddH <sub>2</sub> O to 2L, pH to 8.3
YNB -leu Media (1L)	20g glucose 6.74 YNB 0.7g -leu powder 17.5g agar for plates ddH <sub>2</sub> O to 1L
YPD Media (1L)	10g yeast extract 20g peptone 20g agar for plates 20g dextrose ddH <sub>2</sub> O to 1L
YPD+G418 (300µg/ml)	1L YPD 300mg geneticin (G418) sulfate
YT Media (1L)	10g bacto-tryptone 5g yeast extract 5g NaCl 15g agar for plates 1L ddH <sub>2</sub> O
YT+Ampicillin (50µg/ml)	400ml YT 2ml ampicillin (10mg/ml)

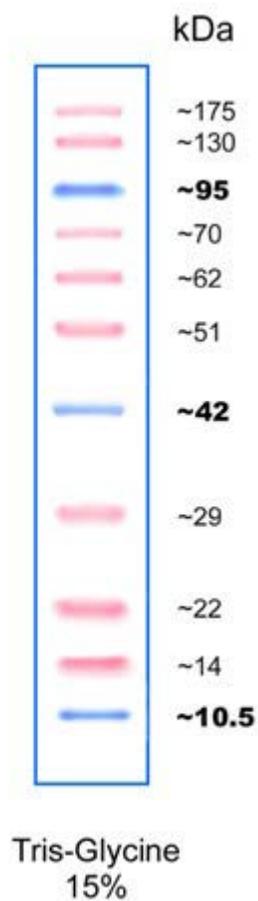
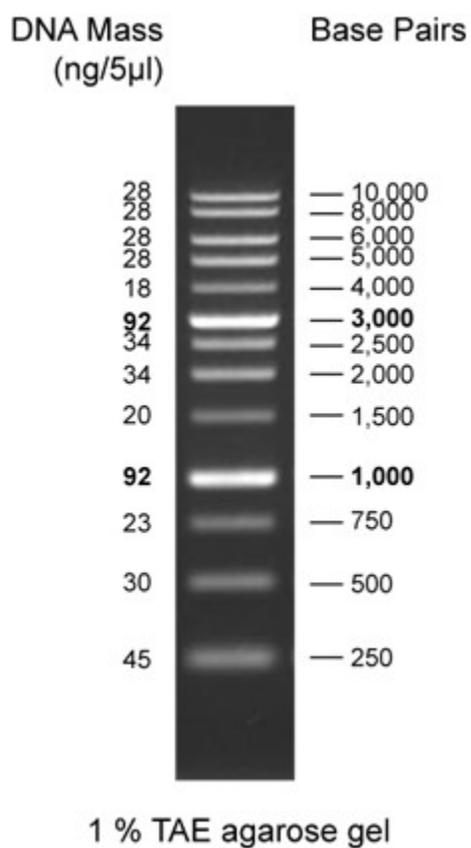
YT+Kanamycin (50µg/ml)	400ml YT 2ml kanamycin (10mg/ml)
------------------------	-------------------------------------

C.6. DNA and protein ladders1

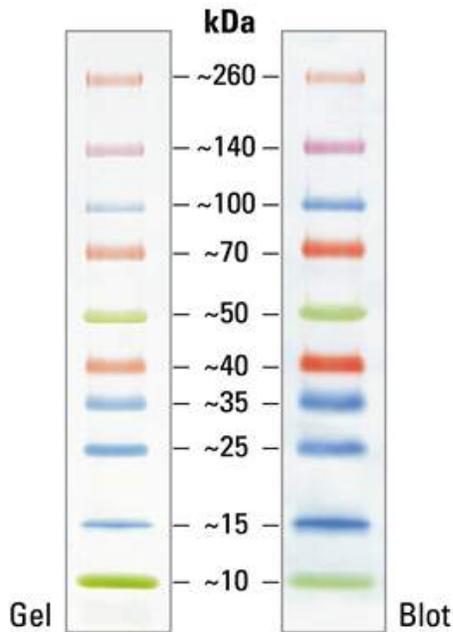
1kb DNA ladder scale (Frogga Bio)

PiNK plus prestained protein ladder scale

(Frogga Bio)



Spectra multicolor broad range protein ladder (Thermo Scientific)



C.7. Sequencing primers

Sequencing primer	Sequence
M13R	5'-CAGGAAACAGCTATGAC-3'
HN111	5'-TATCATCATCATCTTTGTAATCAATATC-3'
M13F	5'-TGTA AACGACGGCCAGT-3'
UF	5'-CTAGCTAAATGTTCCTTTAGCATTTAATTGCAC-3'
H4neoR	5'-TTCAGATTTTGATGCTTCAATAAG-3'
BTU2R	5'-GAGCTAACATGTATGTGAAGAGG-3'
HJ559	5'-CGTTGTAAAACGACGGCCAG-3'

C.8. Homology sequence primers

Med3-FZZ:	UF:	5'-CCCGGTACCTAGCGTACAAGATTCTATTGATAATTTG-3'
	UR:	5'-CCCCTCGAGATAAAAATTCTTTCTTTAAAGCATTATGT-3'
	DF:	5'-CCCGCGGCCGCAACTATTTATTTGTAAATTATTTATACATAC-3'
	DR:	5'-CCCGAGCTCGTATAATTGATCTAATTTCTTCATTTTCAA-3'
Med4-FZZ:	UF:	5'-CCCGGTACCGGTACTAAAAGTTAAAAGATCTAAAAGGAG-3'
	UR:	5'-CCCCTCGAGCTAATATCCCAATCTACTGAACTTGAAGC-3'
	DF:	5'-CCCGCGGCCGCTCCTAAAACATAAATGAACAAAATCTTTATTATCA-3'
	DR:	5'-CCCGAGCTCATATAATCATTGATAATGCTTGCATACTTC-3'

Med17-FZZ: UF: 5'-CCCGGTACCCTTTTTAACATAAATAACCATAAAACCTTTA-3'  
 UR: 5'-CCCCTCGAGTTAAGGAGTTTATGCTGGCATCATCTATTA-3'  
 DF: 5'-CCCGCGGCCGCTGATTAATTCAATTTAATTTGAAATATTTAAAT-3'  
 DR: 5'-GCTTAAATGCTGCTTGAGCTCCTTACATTT-3'

Med20-FZZ: UF: 5'-CCCGGTACCTATTATGTAAGTATGTACCTACCTAGCTAC-3'  
 UR: 5'-CCCCTCGAGTTTTGTAGCTGTTTCAATAAAATAGCTTTT-3'  
 DF: 5'-CCCGCGGCCGCTTTTAATTATTTAATGTGGATCTAAGTTTT-3'  
 DR: 5'-CCCGAGCTCGGCATATTAATAAATAAATAAGATAGAATCTT-3'

Med22-FZZ: UF: 5'-CCCGGTACCTTTTTTGAAAACATAAATAAATATATTGAA-3'  
 UR: 5'-CCCCTCGAGTTAATTTTAAGGATTTATTATTGCTAATA-3'  
 DF: 5'-CCCGCGGCCGCTAGTAAAATCATCATCTTTCATTCATAAAC-3'  
 DR: 5'CCCGAGCTCTTTCTTTAATTAATGTGCAAATTTCAACT-3'

Int2-FZZ: UF: 5'-CCCGGTACCGTGAATGTTACTGCATTTTTCTTCGATTTT-3'  
 UR: 5'-CCCCTCGAGAATGACATATTAATAAGGCCTAAGAGCGAA-3'  
 DF: 5'-CCCGCGGCCGCTGAAAGAAGTTTATAAATAATTTACTTCA-3'  
 DR: 5'-CCCGAGCTCAAGTTGTTTAAATAAAGCAATATAATTAAG-3'

Int4-FZZ: UF: 5'-CCCGGTACCAAAGCTAAAAGTCTGTTTCTGGGCCTCAGC-3'  
 UR: 5'-CCCCTCGAGGATTACCCTAGATTTTATAATTAAGGAAA-3'  
 DF: 5'-CCCGCGGCCGCTGATTAATCCTGGACAAAATAAATTTAAA-3'  
 DR: 5'-CCCGAGCTCCTCAATAAATAAAAACATTGTCTATTGAGG-3'

IntS6-FZZ: UF: 5'-CCCGGTACCCTCAGTCATCTTCACCTGGTCAATTGTCTC-3'  
 UR: 5'-CCCCTCGAGCTTTTTTTCTTAGTAAACAATTTTCAGAGAC-3'  
 DF: 5'-CCCGCGGCCGCGAAGGGTTGATTAATCAAATTAATAATGA-3'  
 DR: 5'-CCCGAGCTCTGATACTTAAATGGATTACAAAAAGTAAAC-3'

Int9-FZZ: UF: 5'-CCCGGTACCGTGAAAACCCTGAGAATCCCTTCTCAGCTT-3'  
 UR: 5'-CCCCTCGAGTTTTAACTCAACTATGTTGTAAAAACAAA-3'  
 DF: 5'-CCCGCGGCCGCTTGAAGAAGATTAATAAAAAATATTATTTG-3'  
 DR: 5'-CCCGAGCTCCGAGAAGCTCTAATAAAAAGCTGTAAATTAAG-3'

Int11-FZZ: UF: 5'-CCCGGTACCGTTTTAATTTTCATCTTTTTTTTTCCAGTTT-3'  
 UR: 5'-CCCCTCGAGTTTATTGTTATATATATTTAAAGCATTGAA-3'  
 DF: 5'-CCCGCGGCCGCTAAAATCAAATGAATTTAATTGCAAGT-3'  
 DR: 5'-CCCGAGCTACATAGAATCCAAAGTCAAAAACTAGCAG-3'

IntS6KO: UF: 5'-CCCGGTACCCTGCCAATAACTGATAAAGTTAGTAAGCA-3'  
 UR: 5'-CCCCTCGAGTCTCTATTTTAAAATAAATTCAATTCCATT-3'  
 DF: 5'-CCCGGATCCGAAGGGTTGATTAATCAAATTAATAATGA-3'  
 DR: 5'-CCCCCGGGGAATAAGAAAAAGAATAAACTACAGAGGAG-3'



**>P.t.**

SFEETVEILYDAAVFSEIDHMRGLSENIIFGQLCPHGTG  
CFELMVNAKNVKEFKLKSSHADKFTQGGEYLAEQSPYDQNOQTPLMLNTPGPGVVSQGFIE NSPYTPYHKSPANFATP  
FGREYTPNSSHCSPFY PNTPLMPNDPYQLSPVGS DSGIQSVQKQANVSDSHSPGSPHYTSH TNSPSPSYR SSERAT  
SGQRSSSISISLSPPSPNYTSSVYNSPLSPTNTGSPRVPTGSPHSPQGSIFTTYS PVYQPGGGTGNQYEQEQ

Serine =  $31/126 \times 100 = 24.65$

Threonine =  $10/126 \times 100 = 7.9\%$

Tyrosine =  $8/126 \times 100 = 6.3\%$

Proline =  $17/126 \times 100 = 13.5\%$

**T.t. vs. P.t.**

Serine:  $|24.6-26.4|/24.6 \times 100 = 7.3\%$

Threonine:  $|7.9-7.4|/7.9 \times 100 = 6.3\%$

Tyrosine:  $|6.3-8.8|/6.3 \times 100 = 39.7\%$

Proline:  $|13.5-12.8|/13.5 \times 100 = 5.2\%$

**>T.t.**

SFEETVDILNTAAIFFEKDDLKGV TENIIFGQNCDIGTG  
CFDLLVDL NKVGEFKTKRQVEQTLEMEDYSENESSRYENTPSHFQTPGPGLVSAYPNSIRTVQHAGSFTPTTPYLNS  
PGYNMSTPISYSNTYNQTRSSQYSPQTGNNSPFVPSPNYSPPSHTPAGSTSPANASPYASSPQYKSSSLSGAHSPSY  
TSPSQVRYNSPSYQNP HASSSSPLDRSKSAYMGSQQSPQYMSSPHYEQRTTPMTRNIKSETSYNPTEERESEESDSD  
QD

Serine =  $39/148 = 26.4\%$

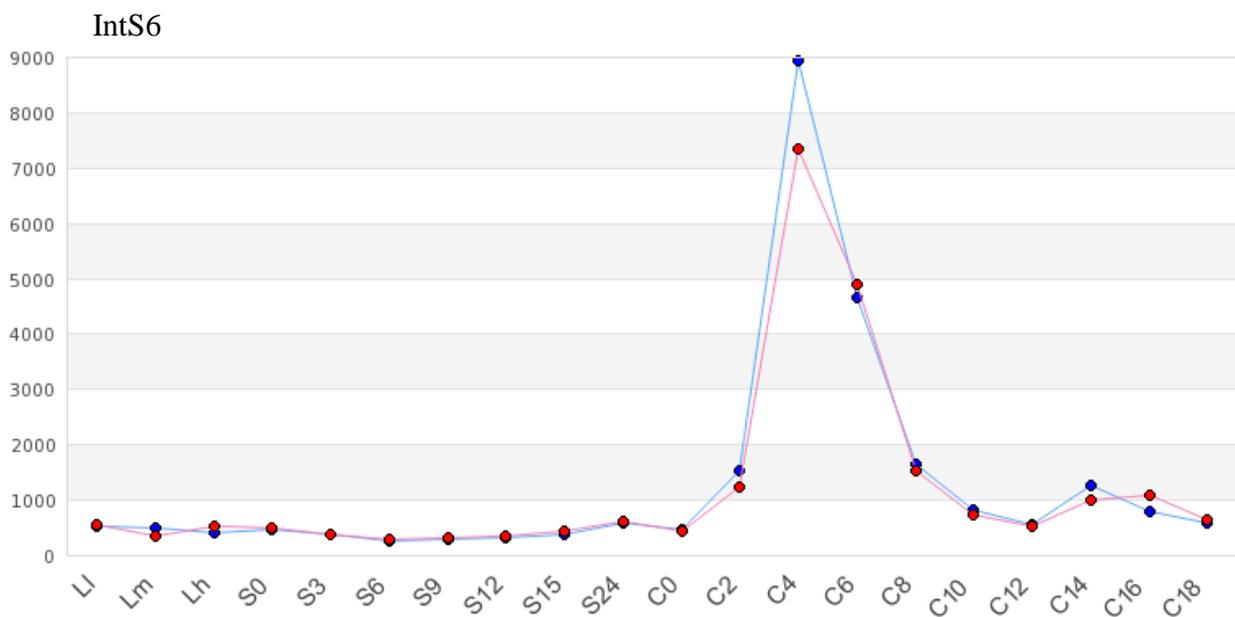
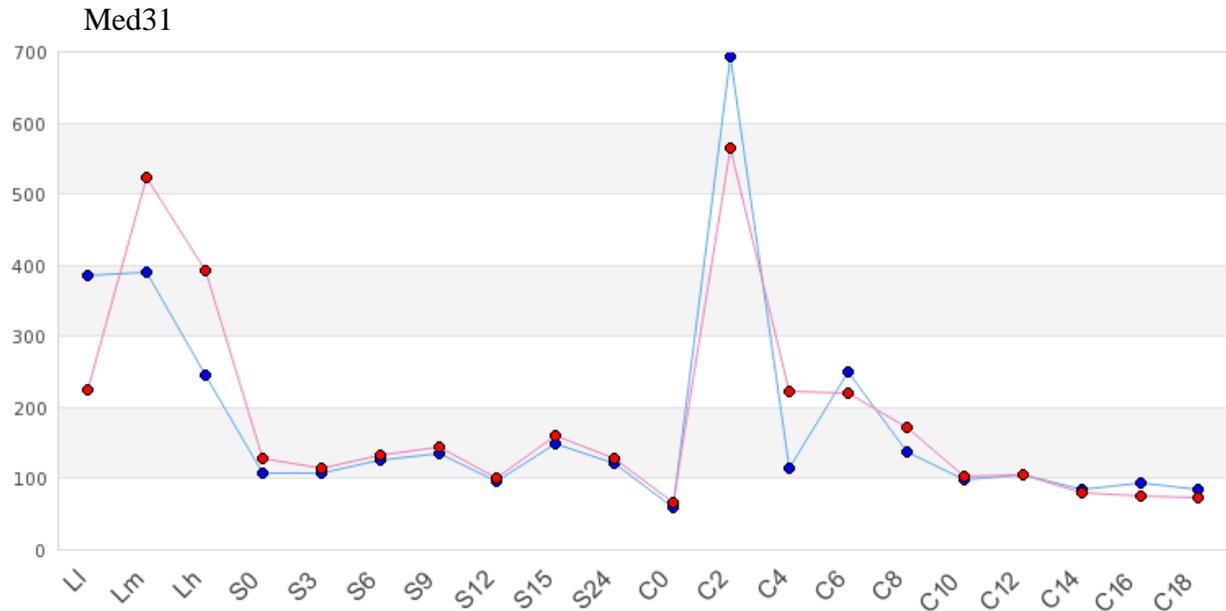
Threonine =  $11/148 \times 100 = 7.4\%$

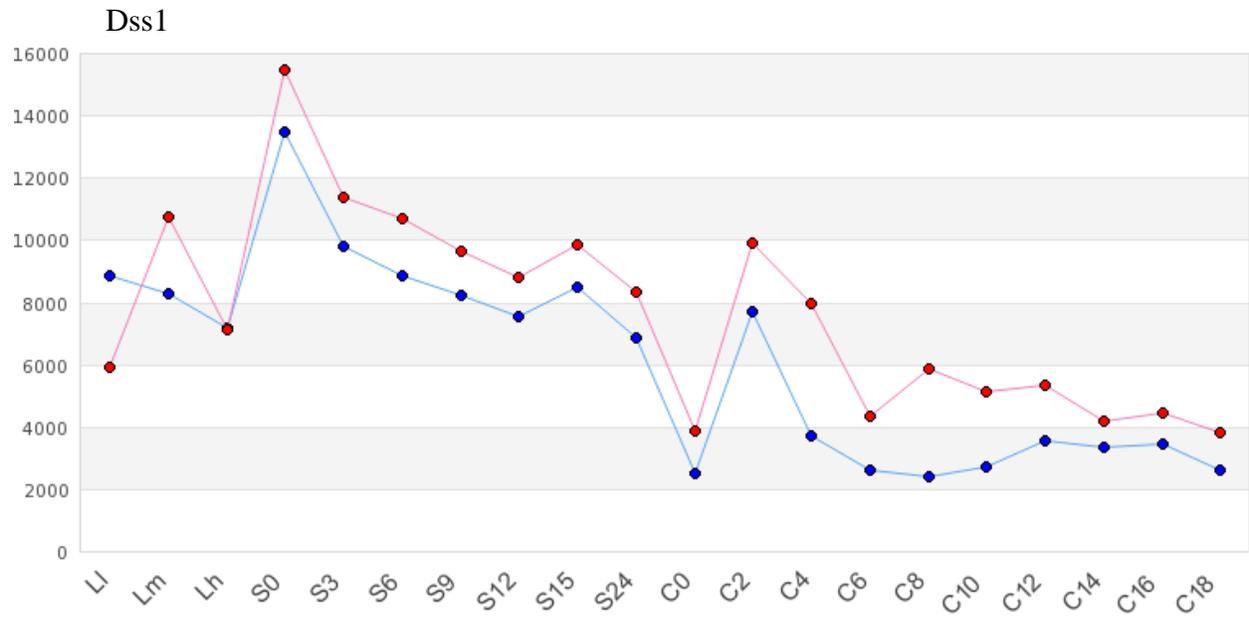
Tyrosine =  $13/148 \times 100 = 8.8\%$

Proline =  $19/148 = 12.8\%$

### C.10. Selected gene expression profiles

Expression profiles compiled from 50 samples completed by Gorovsky, Miao, and Pearlman labs acquired from the Tetrahymena Functional Genomics Database (<http://tfgd.ihb.ac.cn/>). Blue line indicates data from 50 samples prepared by the Gorovsky and Miao labs. Red line indicates data from 50 samples prepared by the Gorovsky and Miao labs plus an additional 10 conjugation samples prepared by the Pearlman lab.





## References

- Akematsu, T., Pearlman, R.E., Endoh, H. (2010). Gigantic macroautophagy in programmed nuclear death of *Tetrahymena thermophila*. *Autophagy* 6, 901-911.
- Akoulitchev, S., Chuikov, S., Reinberg, D. (2000). TFIID is negatively regulated by cdk8-containing mediator complexes. *Nature* 407(6800), 102-106.
- Albrecht, T.R. and Wagner, E.J. (2012). snRNA 3' end formation requires heterodimeric association of integrator subunits. *Mol Cell Biol* 32(6), 1112-1123.
- Allison, L.A., Moyle, M., Shales, M., Ingles, C.J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* 42(2), 599-610.
- Allison, L.A., Wong, J.K., Fitzpatrick, V.D., Moyle, M., Ingles, C.J. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Mol Cell Biol* 8(1), 321-329.
- Armache, K.J., Kettenberger, H., Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci USA* 100(12), 6964-6968.
- Aronica, L., Bednenko, J., Noto, T., Desouza, L.V., Siu, K.W., Loidl, J., Pearlman, R.E., Gorovsky, M.A., Mochizuki, K. (2008). Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev* 22(16), 2228-2241.
- Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M., Kornberg, R.D. (1999). Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 283(5404), 985-987.
- Baillat, D., Hakimi, M.A., Näär, A.M., Shilatifard, A., Cooch, N., Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123(2), 265-276.
- Baldauf, S.L. (2003). The deep roots of eukaryotes. *Science* 300(5626), 1703-1706.
- Bannister, A.J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* 21(3), 381-395.

- Bartkowiak B, Mackellar AL, Greenleaf AL. (2011). Updating the CTD Story: From Tail to Epic. *Genet Res Int* 2011, 623718.
- Bartolomei, M.S., Halden, N.F., Cullen, C.R., Corden, J.L. (1988). Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol Cell Biol* 8(1), 330-339.
- Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7(6), 1513-1523.
- Borggreffe, T., Davis, R., Erdjument-Bromage, H., Tempst, P., Kornberg, R.D. (2002). A complex of the Srb8, -9, -10, and -11 transcriptional regulatory proteins from yeast. *J Biol Chem* 277(46), 44202-44207.
- Borggreffe, T. and Yue, X. (2011). Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Semin Cell Dev Biol* 22(7),759-768.
- Bruns P.J. and Cassidy-Hanley D. (2000). Biolistic transformation of macro- and micronuclei. *Methods Cell Biol* 62, 501-512.
- Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Mol Cell* 36(4), 541-546.
- Cenik, E.S. and Zamore, P.D. (2011). Argonaute proteins. *Curr Biol* 21(12), R446-R449.
- Cervantes, M.D., Hamilton, E.P., Xiong, J., Lawson, M.J., Yuan, D., Hadjithomas, M., Miao, W., Orias, E. (2013). Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLoS Biol* 11(3), e1001518.
- Cervantes, M.D., Xi, X., Vermaak, D., Yao, M.C., Malik, H.S. (2005). The CNA1 histone of the ciliate *Tetrahymena thermophila* is essential for chromosome segregation in the germline micronucleus. *Mol Biol Cell* 17(1), 485-497.
- Chalker, D.L., La Terza, A., Wilson, A., Kroenke, C.D., Yao, M.C. (1999). Flanking regulatory sequences of the *Tetrahymena* R deletion element determine the boundaries of DNA rearrangement. *Mol Cell Biol* 19(8), 5631-5641.

- Chalker, D.L. and Yao, M.C. (1996). Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. *Mol Cell Biol* *16*(7), 3658-3667.
- Chen, J., Ezzeddine, N., Waltenspiel, B., Albrecht, T.R., Warren, W.D., Marzluff, W.F., Wagner, E.J. (2012). An RNAi screen identifies additional members of the *Drosophila* Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA* *18*(12), 2148-2156.
- Chen, J. and Wagner, E.J. (2010). snRNA 3' end formation: the dawn of the Integrator complex. *Biochem Soc Trans* *38*(4), 1082-1087.
- Cheng, C.Y., Vogt, A., Mochizuki, K., Yao, M.C. (2010). A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila*. *Mol Biol Cell* *21*(10), 1753-1762.
- Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* *15*(24), 3319-3329.
- Cho, E.J., Takagi, T., Moore, C.R., Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* *11*(24), 3319-3326.
- Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C., Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* *8*(1), 70-73.
- Collins, K. (2012). Perspectives on the ciliated protozoan *Tetrahymena thermophila*. *Methods Cell Biol* *109*, 1-7.
- Conaway, R.C. and Conaway, J.W. (2011). Origins and activity of the Mediator complex. *Semin Cell Dev Biol* *22*(7), 729-734.
- Conaway, R.C., Sato, S., Tomomori-Sato, C., Yao, T., Conaway, J.W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem Sci* *30*(5), 250-255.
- Cuello, P., Boyd, D.C., Dye, M.J., Proudfoot, N.J., Murphy, S. (1999). Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *EMBO J* *18*, 2867-2877.

- Das, A. and Bellofatto, V. (2009). The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in *Trypanosoma brucei*. *PLoS One* 4(9), e6959.
- de la Mata, M. and Kornblihtt, A.R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol* 13(11), 973-980.
- Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. (2007). The expanding RNA polymerase III transcriptome. *Trends Genet* 23(12), 614-622.
- Dotson, M.R., Yuan, C.X., Roeder, R.G., Myers, L.C., Gustafsson, C.M., Jiang, Y.W., Li, Y., Kornberg, R.D., Asturias, F.J. (2000). Structural organization of yeast and mammalian mediator complexes. *Proc Natl Acad Sci USA* 97(26), 14307-14310.
- Dyballa N. and Metzger S. (2012). Fast and sensitive coomassie staining in quantitative proteomics. *Methods Mol Biol* 893, 47-59.
- Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* 318(5857),1777-1779.
- Egloff, S., O'Reilly, D., Murphy, S. (2008). Expression of human snRNA genes from beginning to end. *Biochem Soc Trans* 36(4), 590-594.
- Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., Murphy, S. (2010). The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* 285(27),20564-20569.
- Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M., Tallon, L.J., Delcher, A.L., Salzberg, S.L., Silva, J.C., Haas, B.J., Majoros, W.H., Farzad, M., Carlton, J.M., Smith, R.K. Jr., Garg, J., Pearlman, R.E., Karrer, K.M., Sun, L., Manning, G., Elde, N.C., Turkewitz, A.P., Asai, D.J., Wilkes, D.E., Wang, Y., Cai, H., Collins, K., Stewart, B.A., Lee, S.R., Wilamowska, K., Weinberg, Z., Ruzzo, W.L., Wloga, D., Gaertig, J., Frankel, J., Tsao, C.C., Gorovsky, M.A., Keeling, P.J., Waller, R.F., Patron, N.J., Cherry, J.M., Stover, N.A., Krieger, C.J., del Toro, C., Ryder, H.F., Williamson, S.C., Barbeau, R.A., Hamilton, E.P., Orias, E. (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 4(9), e286.

- Elliot, A.M. and Hayes, R.E. (1953). Mating types in Tetrahymena. *Biological Bulletin* 105(2), 269-284.
- Epstein, L. M. and Forney, J. D. (1984). Mendelian and non-mendelian mutations affecting surface antigen expression in Paramecium tetraurelia. *Mol Cell Biol* 4, 1583-1590.
- Escher, D., Bodmer-Glavas, M., Barberis, A., Schaffner, W. (2000). Conservation of glutamine-rich transactivation function between yeast and humans. *Mol Cell Biol* 20(8), 2774-2782.
- Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., Werner, M. (2008). Mediator-dependent recruitment of TFIID modules in preinitiation complex. *Mol Cell* 31(3), 337-346.
- Ezzeddine, N., Chen, J., Waltenspiel, B., Burch, B., Albrecht, T., Zhuo, M., Warren, W.D., Marzluff, W.F., Wagner, E.J. (2011). A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation. *Mol Cell Biol* 31(2), 328-341.
- Fan, H.Y., Klein, H.L. (1994). Characterization of mutations that suppress the temperature-sensitive growth of the hpr1 delta mutant of Saccharomyces cerevisiae. *Genetics* 137(4), 945-956.
- Fan, Q. and Yao, M.C. (2000). A long stringent sequence signal for programmed chromosome breakage in Tetrahymena thermophila. *Nucleic Acids Res* 28(4), 895-900.
- Fillingham, J.S. and Pearlman, R.E. (2004). Role of micronucleus-limited DNA in programmed deletion of mse2.9 during macronuclear development of Tetrahymena thermophila. *Eukaryot Cell* 3, 288-301.
- Fillingham, J., Recht, J., Silva, A.C., Suter, B., Emili, A., Stagljar, I., Krogan, N.J., Allis, C.D., Keogh, M.C., Greenblatt, J.F. (2008). Chaperone control of the activity and specificity of the histone H3 acetyltransferase Rtt109. *Mol Cell Biol* 28(13), 4342-4353.
- Fillingham, J.S., Thing, T.A., Vythilingum, N., Keuroghlian, A., Bruno, D., Golding, G.B., Pearlman, R.E. (2004). A non-long terminal repeat retrotransposon family is restricted to the germ line micronucleus of the ciliated protozoan Tetrahymena thermophila. *Eukaryot Cell* 3(1), 157-169.
- Fondell, J.D., Ge, H., Roeder, R.G. (1996). Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc Natl Acad Sci USA* 93(16), 8329-8333.

- Gaertig, J., Gu, L., Hai, B., Gorovsky, M.A. (1994). High frequency vector-mediated transformation and gene replacement in *Tetrahymena*. *Nucleic Acids Res* 22(24), 5391-5398.
- Ghavi-Helm, Y., Michaut, M., Acker, J., Aude, J.C., Thuriaux, P., Werner, M., Soutourina, J. (2008). Genome-wide location analysis reveals a role of TFIIS in RNA polymerase III transcription. *Genes Dev* 22(14), 1934-1947.
- Ghosh, A., Shuman, S., Lima, C.D. (2008). The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Mol Cell* 32(4), 478-490.
- Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., Bentley, D.L. (2009). TFIIF-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol* 29(20), 5455-5464.
- Godiska, R., James, C., Yao, M.C. (1993). A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in *Tetrahymena*. *Genes Dev* 7(12A), 2357-2365.
- Gorovsky, M.A. and Woodard, J. (1969). Studies on the nuclear structure and function in *Tetrahymena pyriformis*. *J Cell Biol* 42, 673-682.
- Grayson, M.A. (2011). John Bennett Fenn: a curious road to the prize. *J Am Soc Mass Spectrom* 22(8), 1301-1308.
- Greider, C.W. and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* 43, 405-413.
- Guglielmi, B., Soutourina, J., Esnault, C., Werner, M. (2007). TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc Natl Acad Sci USA* 104(41), 16062-16067.
- Hamatake, R.K., Dykstra, C.C., Sugino, A. (1989). Presynapsis and synapsis of DNA promoted by the STP alpha and single-stranded DNA-binding proteins from *Saccharomyces cerevisiae*. *J Biol Chem* 264(22), 13336-13342.
- Han, S.M., Lee, T.H., Mun, J.Y., Kim, M.J., Kritikou, E.A., Lee, S.J., Han, S.S., Hengartner, M.O., Koo, H.S. (2006). Deleted in cancer 1 (DICE1) is an essential protein controlling

- the topology of the inner mitochondrial membrane in *C. elegans*. *Development* *133*(18), 3597-3606.
- Harper, S. and Speicher, W. (2011). Purification of proteins fused to glutathione S-transferase. *Methods Mol Biol.* *691*, 259-280.
- Hengartner, C.J., Thompson, C.M., Zhang, J., Chao, D.M., Liao, S.M., Koleske, A.J., Okamura, S., Young, R.A. (1995). Association of an activator with an RNA polymerase II holoenzyme. *Genes Dev* *9*(8), 897-910.
- Hentges, K.E. (2011). Mediator complex proteins are required for diverse developmental processes. *Semin Cell Dev Biol* *22*(7), 769-775.
- Hernandez, N. (1985). Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J* *4*(7), 1827-1837.
- Ho, C.S., Lam, C.W., Chan, M.H., Cheung, R.C., Law, L.K., Lit, L.C., Ng, K.F., Suen, M.W., Tai, H.L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* *24*(1), 3-12.
- Horowitz, S. and Gorovsky, M.A. (1985). An unusual genetic code in nuclear genes of *Tetrahymena*. *Proc Natl Acad Sci USA* *82*(8), 2452-2455.
- Hsin, J.P. and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* *26*(19), 2119-2137.
- Hsin, J.P., Sheth, A., Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* *334*(6056), 683-686.
- Igarashi, K. and Kashiwagi, K. (2000). Polyamines: mysterious modulators of cellular functions. *Biochem Biophys Res Commun* *271*(3), 559-564.
- Imasaki, T., Calero, G., Cai, G., Tsai, K.L., Yamada, K., Cardelli, F., Erdjument-Bromage, H., Tempst, P., Berger, I., Kornberg, G.L., Asturias, F.J., Kornberg, R.D., Takagi, Y. (2011). Architecture of the Mediator head module. *Nature* *475*(7355), 240-243.
- International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature* *409*(6822), 860-921.

- Johnson, K.M., Wang, J., Smallwood, A., Arayata, C., Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes Dev* *16(14)*, 1852-1863.
- Karrer, K.M. (2012). Nuclear dualism. *Methods Cell Biol* *109*, 29-52.
- Karrer, K., Stein-Gavens, S., Allitto, B.A. (1984). Micronucleus-specific DNA sequences in an amiconucleate mutant of *Tetrahymena*. *Dev Biol* *105(1)*, 121-129.
- Kean, M.J., Couzens, A.L., Gingras, A.C. (2012). Mass spectrometry approaches to study mammalian kinase and phosphatase associated proteins. *Methods* *57(4)*, 400-408.
- Kelly, W.G., Dahmus, M.E., Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* *268(14)*, 10416-10424.
- Keogh, M.C., Mennella, T.A., Sawa, C., Berthelet, S., Krogan, N.J., Wolek, A., Podolny, V., Carpenter, L.R., Greenblatt, J.F., Baetz, K., Buratowski, S. (2006). The *Saccharomyces cerevisiae* histone H2A variant Htz1 is acetylated by NuA4. *Genes Dev* *20(6)*, 660-665.
- Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* *77(4)*, 599-608.
- Komarnitsky, P., Cho, E.J., Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* *14(19)*, 2452-2460.
- Koschubs, T., Seizl, M., Larivière, L., Kurth, F., Baumli, S., Martin, D.E., Cramer, P. (2009). Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *EMBO J* *28(1)*, 69-80.
- Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., Hampsey, M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* *14(3)*, 387-394.
- Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrin-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K.,

- Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., Greenblatt, J.F. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* *440*(7084), 637-643.
- Krogan, N.J., Lam, M.H., Fillingham, J., Keogh, M.C., Gebbia, M., Li, J., Datta, N., Cagney, G., Buratowski, S., Emili, A., Greenblatt, J.F. (2004). Proteasome involvement in the repair of DNA double-strand breaks. *Mol Cell* *16*(6), 1027-1034.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* *31*(1), 147-157.
- Kuras, L., Borggrefe, T., Kornberg, R.D. (2003). Association of the Mediator complex with enhancers of active genes. *Proc Natl Acad Sci USA* *100*(24), 13887-13891.
- Kurth, H.M. and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* *15*(4), 675-685.
- LaFountain, J.R. Jr. and Davidson, L.A. (1979). An analysis of spindle ultrastructure during prometaphase and metaphase of micronuclear division in *Tetrahymena*. *Chromosoma* *75*, 293-308.
- Landry, J. and Sternglanz, R. (2009). Yeast *Fms1* is a FAD-utilizing polyamine oxidase. *Biochem Biophys Res Commun* *303*(3), 771-776.
- Laybourn, P.J. and Dahmus, M.E. (1989). Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o. *J Biol Chem* *264*(12), 6693-6698.
- Laybourn, P.J. and Dahmus, M.E. (1990). Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J Biol Chem* *265*(22), 13165-13173.
- Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* *23*(20), 4051-4060.

- Li, H., Zhang, Z., Wang, B., Zhang, J., Zhao, Y., Jin, Y. (2007). Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. *Mol Cell Biol* 27(15), 5296-5305.
- Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., Hahn, S. (2004a). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol Cell Biol* 24(4), 1721-1735.
- Liu, Y., Mochizuki, K., Gorovsky, M.A. (2004b). Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. *Proc Natl Acad Sci USA* 101(6), 1679-1684.
- Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. *Genes Dev* 21(12), 1530-1545.
- Longtine, M.S., McKenzie, A. 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* 14(10), 953-961.
- Lopes da Rosa, J., Boyartchuk, V.L., Zhu, L.J., Kaufman, P.D. (2010). Histone acetyltransferase Rtt109 is required for *Candida albicans* pathogenesis. *Proc Natl Acad Sci USA* 107(4), 1594-1599.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648), 251-260.
- Lynn, D.H. and Doerder, F.P. (2012). The life and times of *Tetrahymena*. *Methods Cell Biol* 109, 9-27.
- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444(7121), 953-956.
- March, R.E. (1997). An introduction to quadrupole ion trap mass spectrometry. 32, 351-369.
- Marston, N.J., Richards, W.J., Hughes, D., Bertwistle, D., Marshall, C.J., Ashworth, A. (1999). Interaction between the product of the breast cancer susceptibility gene BRCA2

- andDSS1, a protein functionally conserved from yeast to mammals. *Mol Cell Biol* *19*(7), 4633-4642.
- Matheny, P.B., Liu, Y.J., Ammirati, J.F., Hall, B.D. (2002). Using RPB1 sequences to improve phylogenetic inference among mushrooms (Inocybe, Agaricales). *Am J Bot* *89*(4), 688-698.
- Mathur, S., Vyas, S., Kapoor, S., Tyagi, A.K. (2011). The Mediator complex in plants: structure, phylogeny, and expression profiling of representative genes in a dicot (*Arabidopsis*) and a monocot (rice) during reproduction and abiotic stress. *Plant Physiol* *157*(4), 1609-1627.
- Mayer, A., Heidemann, M., Lidschreiber, M., Schrieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* *336*(6089), 1723-1725.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* *385*(6614), 357-361.
- Medlin, J.E., Uguen, P., Taylor, A., Bentley, D.L., Murphy, S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J* *22*(4), 925-934.
- Mehta, R. and Champney, W.S. (2003). Neomycin and paromomycin inhibit 30S ribosomal subunit assembly in *Staphylococcus aureus*. *Curr Microbiol* *47*(3), 237-243.
- Merriam, E.V. and Bruns, P.J.(1988). Phenotypic assortment in *Tetrahymena thermophila*: assortment kinetics of antibiotic-resistance markers, tsA, death, and the highly amplified rDNA locus. *Genetics* *120*(2), 389-395.
- Mochizuki, K. (2010). DNA rearrangements directed by non-coding RNAs in ciliates. *Wiley Interdiscip Rev RNA* *1*(3), 376-387.
- Mochizuki, K. (2012). Developmentally programmed, RNA-directed genome rearrangement in *Tetrahymena*. *Dev Growth Differ* *54*(1), 108-119.
- Mochizuki, K., Fine, N.A., Fujisawa, T., Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* *110*(6), 689-699.

- Mochizuki, K. and Gorovsky, M.A. (2004). Small RNAs in genome rearrangement in Tetrahymena. *Curr Opin Genet Dev* 14(2), 181-187.
- Mochizuki K, Gorovsky MA. (2005). A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev* 19(1), 77-89.
- Mochizuki, K., Novatchkova, M., Loidl, J. (2008). DNA double-strand breaks, but not crossovers, are required for the reorganization of meiotic nuclei in Tetrahymena. *J Cell Sci* 121(13), 2148-2158.
- Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell* 34(2), 168-178.
- Mumby, M.C. and Walter, G. (1993) Protein serine/threonine phosphatases: structure, regulation, and functions in cell growth. *Physiol Rev* 73(4), 673-699.
- Myers, L.C. and Kornberg, R.D. (2000). Mediator of transcriptional regulation. *Annu Rev Biochem* 69, 729-749.
- Noto, T., Kurth, H.M., Kataoka, K., Aronica, L., Desouza, L.V., Siu, K.W., Pearlman, R.E., Gorovsky, M.A. & Mochizuki, K. (2010). The tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. *Cell* 140(5), 692-703.
- Orias, E. and Newby, C.J. (1975). Macronuclear genetics of Tetrahymena. II. Macronuclear location of somatic mutations to cycloheximide resistance. *Genetics* 80(2), 251-262.
- Park, M.H., Nishimura, K., Zanelli, C.F., Valentini, S.R. (2010). Functional significance of eIF5A and its hypusine modification in eukaryotes. *Amino Acids* 38(2), 491-500.
- Petukhova GV, Romanienko PJ, Camerini-Otero RD. (2003). The Hop2 protein has a direct role in promoting interhomolog interactions during mouse meiosis. *Dev Cell* 5(6), 927-936.
- Phatnani, H.P. and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* 20(21), 2922-2936.

- Pollard, K.J., Samuels, M.L., Crowley, K.A., Hansen, J.C., Peterson, C.L. (1999). Functional interaction between GCN5 and polyamines: a new role for core histone acetylation *EMBO J* 18(20), 5622-5633.
- Radovani, E., Cadorin, M., Shams, T., El-Rass, S., Karsou, A.R., Kim, H.S., Kurat, C.F., Keogh, M.C., Greenblatt JF, Fillingham JS. (2013). The carboxyl terminus of Rtt109 functions in chaperone control of histone acetylation. *Eukaryot Cell* 12(5), 654-664.
- Ray, C. Jr. (1956). Preparation of chromosomes of *Tetrahymena pyriformis* for photomicrography. *Stain Technol* 31(6), 271-274.
- Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., Buratowski, S. (2000). Kin28, the TFIIF-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol Cell Biol* 20(1), 104-112.
- Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 21(9), 327-335.
- Romanienko, P.J. and Camerini-Otero, R.D. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol Cell* 6(5), 975-987.
- Russel, J. and Zomerdijk, J.C. (2006). The RNA polymerase I transcription machinery. *Biochem Soc Symp* (73), 203-216.
- Schiestl, R.H. and Gietz, R.D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* 16(5-6), 339-346.
- Schnapp, G., Graveley, B.R., Grummt, I. (1996). TFIIS binds to mouse RNA polymerase I and stimulates transcript elongation and hydrolytic cleavage of nascent rRNA. *Mol Gen Genet* 252(4), 412-419.
- Selth, L. and Svejstrup, J.Q. (2007). Vps75, a new yeast member of the NAP histone chaperone family. *J Biol Chem* 282(17), 12358-12362.
- Shahbazian, M.D. and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem* 76, 75-100.
- Sidow, A. and Thomas, W.K. (1994). A molecular evolutionary framework for eukaryotic model organisms. *Curr Biol* 4(7), 596-603.

- Sims, R.J. 3rd., Rojas, L.A., Beck, D., Bonasio, R., Schuller, R., Drury W.J., 3rd., Eick, D., Reinberg, D. (2011). The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* 332(6025), 99-102.
- Sone, T., Saeki, Y., Toh-e, A., Yokosawa, H. (2004). Sem1p is a novel subunit of the 26S proteasome from *Saccharomyces cerevisiae*. *J Biol Chem* 279(27), 28807-28816.
- Spaeth, J.M., Kim, N.H., Boyer, T.G. (2011). Mediator and human disease. *Semin Cell Dev Biol* 22(7), 776-787
- Stover, N.A., Krieger, C.J., Binkley, G., Dong, Q., Fisk, D., Nash, R., Sethuraman, A., Weng, S., Cherry, M. (2006). Tetrahymena Genome Database (GDB): a new genomic resource for tetrahymena thermophila research. *Nucleic Acids Res* 34, D500-D503.
- Stump, A.D. and Ostrozhynska, K. (2013). Selective constraint and evolution of the RNA polymerase II C-Terminal Domain. *Transcription* 4(2), 77-86.
- Sugino, A., Nitiss, J., Resnick, M.A. (1988). ATP-independent DNA strand transfer catalyzed by protein(s) from meiotic cells of the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 85(11), 3683-3687.
- Sun X, Zhang Y, Cho H, Rickert P, Lees E, Lane W, Reinberg D. (1998). NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol Cell* 2(2), 213-222.
- Sydow, J.F. and Kramer, P. (2009). RNA polymerase fidelity and transcriptional proofreading. *Curr Opin Struct Biol.* 19(6), 732-739.
- Tabor, C.W. and Tabor, H. (1984). Polyamines. *Annu Rev Biochem* 53, 749-90.
- Taverna, S.D., Coyne, R.S. & Allis, C.D. 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. *Cell* 110(6), 701-711.
- Taylor, S.S. and Kornev, A.P. (2011). Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem Sci* 36(2), 65-77.
- Uguen, P. and Murphy, S. (2003). The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *EMBO J* 22(17), 4544-4554.

- Vavra, K.J., Allis, C.D., and Gorovsky, M.A. (1982). Regulation of histone acetylation in *Tetrahymena* macro- and micronuclei. *J Biol Chem* 257, 2591-2598.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297(5588), 1833-1837.
- Werner-Allen, J.W., Lee, C.J., Liu, P., Nicely, N.I., Wang, S., Greenleaf, A.L., Zhou, P. (2011). cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *J Biol Chem* 286(7), 5717-5726.
- West, M.L. and Corden, J.L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140(4), 1223-1233.
- Whittaker, C.A. and Hynes, R.O. (2002). Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* 13(10), 3369-3387.
- Wieland, I., Arden, K.C., Michels, D., Klein-Hitpass, L., Böhm, M., Viars, C.S., Weidle, U.H. (1999). Isolation of DICE1: a gene frequently affected by LOH and downregulated in lung carcinomas. *Oncogene* 18(32), 4530-4537.
- Wierzbicki, A.T., Hagg, J.R., Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135, 635-648.
- Wilcox CB, Rossetini A, Hanes SD. (2004). Genetic interactions with C-terminal domain (CTD) kinases and the CTD of RNA Pol II suggest a role for ESS1 in transcription initiation and elongation in *Saccharomyces cerevisiae*. *Genetics* 167(1), 93-105.
- Will, C.L. and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3(7).
- Wind, M. and Reines, D. (2000). Transcription elongation factor SII. *Bioessays* 22(4), 327-336.
- Wolff, E.C., Kang, K.R., Kim, Y.S., Park, M.H. (2007). Posttranslational synthesis of hypusine: evolutionary progression and specificity of the hypusine modification. *Amino Acids* 33(2), 341-350.

- Xiao, H., Jeang, K.T. (1998). Glutamine-rich domains activate transcription in yeast *Saccharomyces cerevisiae*. *J Biol Chem* 273(36), 22873-22876.
- Xiong J, Yuan D, Fillingham JS, Garg J, Lu X, Chang Y, Liu Y, Fu C, Pearlman RE, Miao W. (2011). Gene network landscape of the ciliate *Tetrahymena thermophila*. *PLoS One* 6(5), e20124.
- Yang, X.C., Sullivan, K.D., Marzluff, W.F., Dominski, Z. (2009). Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Mol Cell Biol* 29(1), 31-42.
- Yao, M.C. and Chao, J.L. (2005). RNA-guided DNA deletion in *Tetrahymena*: an RNAi-based mechanism for programmed genome rearrangements. *Annu Rev Genet* 39, 537-359.
- Yao, M.C., Choi, J., Yokoyama, S., Austerberry, C.F., Yao, C.H. (1984). DNA elimination in *Tetrahymena*: a developmental process involving extensive breakage and rejoining of DNA at defined sites. *Cell* 36(2), 433-440.
- Yao, M.C., Yao, C.H., Monks, B. (1990). The controlling sequence for site-specific chromosome breakage in *Tetrahymena*. *Cell* 63(4), 763-772.
- Young, R.A. (1991). RNA polymerase II. *Annu Rev of Biochem* 60, 689-715.
- Zanelli, C.F., Maragno, A.L., Gregio, A.P., Komili, S., Pandolfi, J.R., Mestriner, C.A., Lustrì, W.R., Valentini, S.R. (2006). eIF5A binds to translational machinery components and affects translation in yeast. *Biochem Biophys Res Commun* 348(4), 1358-1366.
- Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., Greenleaf, A.L. (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. *Proc Natl Acad Sci USA* 85(11), 3698-3702.
- Zhang, L., Maul, R.S., Rao, J., Apple, S., Seligson, D., Sartippour, M., Rubio, R., Brooks, M.N. (2004). Expression pattern of the novel gene EG-1 in cancer. *Clin Cancer Res* 10(10), 3504-3508.
- Zhou, M., Halanski, M.A., Radonovich, M.F., Kashanchi, F., Peng, J., Price, D.H., Brady, J.N. (2000). Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. *Mol Cell Biol* 20(14), 5077-5086.