MECHANISM AND TARGETS OF SEN1 HELICASE IN TERMINATION OF NON-CODING RNA SYNTHESIS IN YEAST

by

Xin Chen

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the UNIVERSITY OF WISCONSIN-MADISON

2015

Date of final oral examination: 06/03/2015

The dissertation is approved by the following members of the Final Oral Committee:

David A. Brow, Professor, Department of Biomolecular Chemistry

Catherine A. Fox, Professor, Department of Biomolecular Chemistry

James L. Keck, Professor, Department of Biomolecular Chemistry

Deane F. Mosher, Professor, Department of Biomolecular Chemistry

David J. Eide, Professor, Department of Nutritional Sciences

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude for people who helped me along the way.

First and foremost, I would like to thank my advisor, Dr. David Brow, for offering me the precious opportunity of pursuing science in his group, and guiding me all the way. His trust, encouragement, patience and wisdom have been invaluable not only for my career in science but also my life. I am particularly grateful to the efforts he has devoted to improving my writing and presentation skills.

I also want to thank all the current and former members in my thesis committing, Dr. Catherine Fox, Dr. James Keck, Dr. Deane Mosher, Dr. David Eide and Dr. David Wassarman for their support and insightful advice during my graduate training.

I would like to thank the past and current members of the Brow lab, including Dr. Eric Montemayor, Dr. Steve Tumasz-Martin, Dr. Jordan Burke, Dr. Kaitlin Sundling, Ulrika Müller, Allyson Dill, Erik Anderson, Sean Hinds, for being the great collaborators and contributing to the excellent work environment.

I would like to acknowledge the members of Dr. Samuel Butcher and Dr. Aaron Hoskins' labs for organizing the joint lab meetings that have broadened my horizons in the amazing world of RNA. I would also like to thank the staff members in the IPiB program and the Department of Biomolecular Chemistry for their hard work and dedication.

I am especially grateful to my family and friends. Their encouragement and support help me conquer all the challenges and difficulties along the way.

Lastly, I would like to dedicate this dissertation to my parents for their understanding and unconditional love.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

MECHANISM AND TARGETS OF SEN1 HELICASE IN TERMINATION OF NON-CODING RNA SYNTHESIS IN YEAST

Xin Chen

Under the supervision of Professor David A. Brow

at the University of Wisconsin-Madison

ABSTRACT

In the yeast S. *cerevisiae*, the essential gene *SEN1* encodes a superfamily I helicase that is required for terminating transcription of many non-coding RNA genes by RNA polymerase II. Sen1 also regulates transcription of some protein-coding genes by mechanisms such as attenuation and antisense transcription. Using a yeast genetic system, I showed that the helicase domain plus a flanking nuclear localization sequence is the minimal essential region of Sen1. Guided by the Upf1 structure model, I found the termination defect caused by the *sen1-E1597K* mutation is disruption of an intradomain salt bridge in the helicase domain. Mutations in the human ortholog of Sen1, Senataxin, co-segregate with the human disease Ataxia oculomotor apraxia (AOA) 2. I characterized 13 AOA2 missense mutations that are located in the helicase domain in the context of Sen1, and identified termination defects as a common feature of those disease mutations.

Using hypomorphic mutations, I investigated the influences of Sen1, Nrd1, Nab3, Ssu72, Rpb11 and Hrp1 on Sen1-dependent termination. Although these six proteins are all involved in Sen1-dependent termination, the genome-wide effects caused by their mutations are surprisingly diverse. The substitution in Sen1 induces the broadest and strongest termination defects, while the others only affect termination of variable subsets of Sen1's targets. I confirmed that these factors participate in transcription termination of sn/snoRNA genes, but to different degrees, and

further showed that the RNA-binding abilities of Nrd1 and Nab3 are dispensable for Sen1-dependent termination of some snoRNA genes. I also found that Sen1 may have a regulatory role in repressing meiotic genes in vegetative cells. The findings of my transcriptome analysis suggest that Sen1-dependent termination is under complex, combinatorial control.

CHAPTER 1: INTRODUCTION

In eukaryotic cells, RNA polymerase II (Pol II) transcribes genes to synthesize all the mRNAs and many non-coding RNAs, including small nuclear (sn) and small nucleolar (sno) RNAs, micro (mi) RNAs, and a number of unstable/uncharacterized RNAs. There are three major essential stages in the transcription cycle, initiation, elongation and termination, and this dissertation will be mostly focused on the termination stage.

Transcription termination has vital functions for gene expression. It determines the end of transcription, therefore defines the boundaries of transcription units and the genetic messages carried by the RNAs. Transcription termination factors can interact with RNA processing and degradation enzymes, and the selectivity of their interacting partners determines the cellular fate of the transcript (PORRUA and LIBRI 2015). In the yeast *Saccharomyces cerevisiae*, transcription termination by Pol II uses at least two pathways (Figure 1.1), one that is coupled to cleavage and polyadenylation of the nascent transcript (the poly(A)-dependent pathway) and one that involves the activity of the RNA/DNA helicase Sen1 (the Sen1-dependent pathway) (KIM *et al.* 2006; KUEHNER *et al.* 2011; PORRUA and LIBRI 2015).

1.1. Two major pathways that terminate Pol II transcription

The poly(A)-dependent transcription termination pathway

The poly(A)-dependent pathway terminates transcription of Pol II downstream of most protein-coding genes, featured by cleavage and polyadenylation of the nascent transcript. In brief, after the poly(A) site is transcribed, the nascent transcript is cleaved by the endoribonuclease Ysh1. The 5' cleavage product is polyadenylated by poly(A) polymerase Pap1 to form the mRNA poly(A) tail. The 3' cleavage product is degraded by the 5' to 3' exoribonuclease Rat1 (Xrn2 in human). How Pol II is released after cleavage is unclear, and two major models have been proposed for this process. In the "allosteric" model, the conformational

change of Pol II by loss of the elongation factors and/or association of the polyadenylation/termination factors occurs upon transcription of the poly(A) signal, and may lead to dissociation of the elongation complex (LOGAN *et al.* 1987; KIM *et al.* 2004a). The other one is the "torpedo" model, where Rat1/Xrn2 is recruited to "chew" the transcript to catch up with Pol II and promotes destabilization of the elongation complex (KIM *et al.* 2004b; WEST *et al.* 2004). Both models are supported by experimental data, so a combination of mechanisms may be utilized for poly(A)-dependent termination (RICHARD and MANLEY 2009).

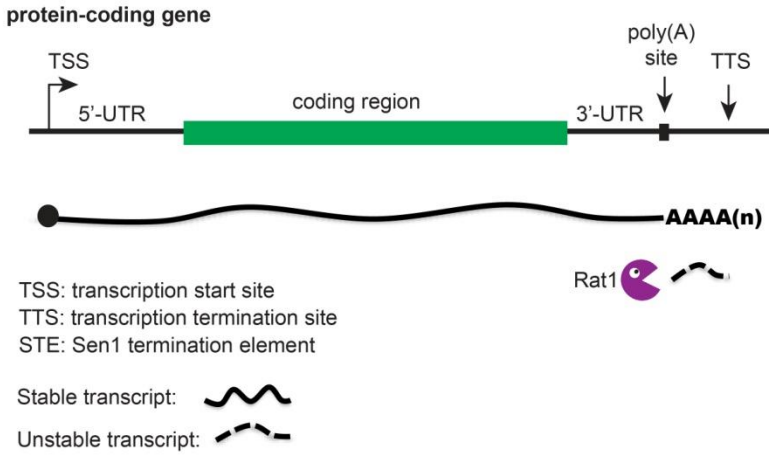
The Sen1-dependent transcription termination pathway

The Sen1-dependent transcription termination pathway is much less studied compared to the poly(A)-dependent pathway. This pathway was first identified in the yeast *S. cerevisiae* and is responsible for transcription termination of many short, non-coding RNA genes, including sn/snoRNAs, CUTs (Cryptic Unstable Transcripts), SUTs (Stable Uncharacterized Transcripts) and NUTs (Nrd1-Unterminated Transcripts) (WINEY and CULBERTSON 1988; STEINMETZ and BROW 1996; RASMUSSEN and CULBERTSON 1998; STEINMETZ *et al.* 2001; MARQUARDT *et al.* 2011; SCHULZ *et al.* 2013). The Sen1 pathway also regulates gene expression by mechanisms including attenuation, transcription interference and antisense transcription (ARIGO *et al.* 2006a; STEINMETZ *et al.* 2006b; JENKS *et al.* 2008; KUEHNER and BROW 2008; CREAMER *et al.* 2011; SCHULZ *et al.* 2013) (Figure 1.1). Moreover, some short mRNA genes, such as *CYC1*, may have hybrid terminators that require factors from not only the poly(A) pathway but also the Sen1 pathway (STEINMETZ *et al.* 2006b).

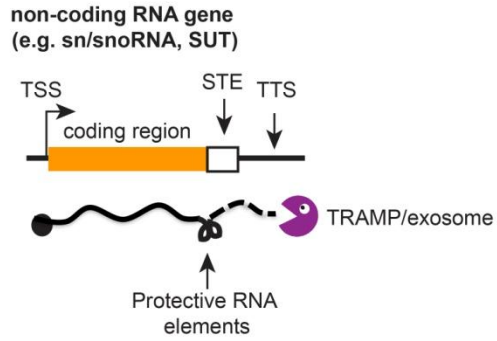
The core termination complex of the Sen1 pathway is distinct from the poly(A) pathway, including two RNA-binding proteins Nrd1 and Nab3, and the helicase Sen1 (Figure 1.2) (STEINMETZ and BROW 1996; STEINMETZ and BROW 1998; CONRAD *et al.* 2000; STEINMETZ *et*

FIGURE 1.1 The two major Pol II-mediated transcription termination pathways and their 3' end formation in yeast. A) Most protein-coding genes utilize the poly(A)-dependent pathway. After cleavage, the 5' cleavage product is polyadenylated for the mature mRNA, and the 3' cleavage product is degraded by the exonuclease Rat1 from 5' to 3'. **B**) and **C**) represent the Sen1-dependent pathway. **B**) For stable non-coding RNAs, including sn/snoRNAs and SUTs, the Sen1-termination elements on the nascent transcript are recognized by Nrd1 and Nab3, which recruit Sen1 to elicit transcription termination. The TRAMP complex and nuclear exosome trim the 3' end of the nascent transcript until reaching a protective RNA element, which is a protein-bound state and/or secondary structure that stops exosome degradation. **C**) Unstable non-coding RNAs, such as CUTs, are terminated and processed similarly as stable ones, but without the protective RNA element, the nascent transcript is completely degraded by the TRAMP complex and nuclear exosome.

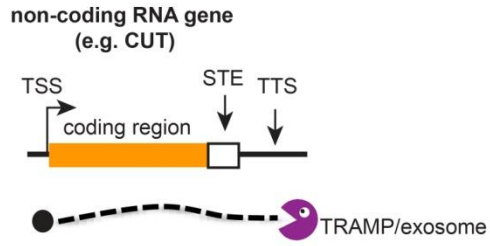
A


protein-coding gene
TSS
5'-UTR
coding region
3'-UTR
poly(A) site
TTS
AAAA(n)
Rat1
TSS: transcription start site
TTS: transcription termination site
STE: Sen1 termination element
Stable transcript:
Unstable transcript:

B


non-coding RNA gene
(e.g. sn/snoRNA, SUT)
TSS
STE
TTS
coding region
TRAMP/exosome
Protective RNA elements

C


non-coding RNA gene
(e.g. CUT)
TSS
STE
TTS
coding region
TRAMP/exosome

al. 2001; CARROLL *et al.* 2007). The mechanism by which this complex terminates transcription has not been fully understood. One prevailing model is that, upon transcription of the Sen1 termination element, Nrd1 and Nab3 recognize their binding motifs in the nascent transcript (Nab3: UCUUG, Nrd1: U/AGUAAA) and recruit Sen1 to terminate transcription (STEINMETZ and BROW 1998; CARROLL *et al.* 2007; PORRUA *et al.* 2012). The helicase Sen1 may function analogously to bacterial Rho helicase, translocating 5' to 3' along the nascent transcript until colliding with paused Pol II and terminating its elongation (STEINMETZ and BROW 1996; PORRUA and LIBRI 2013). Another model is that Sen1 terminates transcription by resolving RNA:DNA hybrids, called R loops. Sen1 unwinds R loops formed between the nascent transcript and the template strand, thereby allowing recognition of termination elements in the nascent transcript by RNA-binding proteins (MISCHO *et al.* 2011). In fact, a unified model can be proposed where Sen1 functions on both resolving R loops and promoting Pol II release for termination.

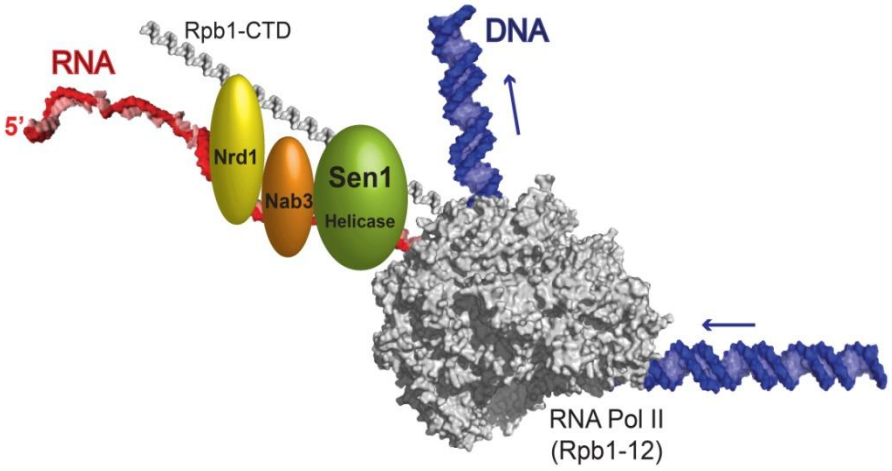
Another distinctive feature of the Sen1-terminated transcripts from the poly(A) pathway is that the Sen1-terminated transcripts are targeted by the nuclear exosome, which trims them for the 3' ends of mature sn/snoRNAs or degrades the unstable ones such as CUTs completely (ALLMANG *et al.* 1999; ARIGO *et al.* 2006b; KIM *et al.* 2006; THIEBAUT *et al.* 2006). The TRAMP complex (Trf4-Air2-Mtr4) is required to catalyze polyadenylation of the transcript by Trf4, and to stimulate the nuclear exosome for RNA processing and degradation (LACAVA *et al.* 2005; WYERS *et al.* 2005).

1.2. Key players in the Sen1-dependent transcription termination pathway

The RNA/DNA helicase Sen1

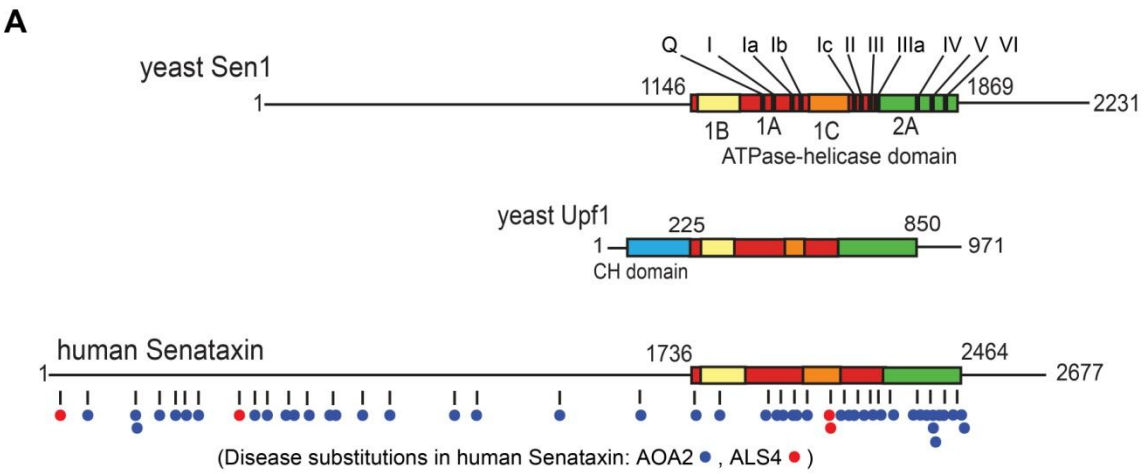
S. cerevisiae Sen1 is a 252 kDa superfamily 1 helicase encoded by the essential *SEN1* gene. The helicase domain of Sen1 is located in the C-terminal half of the protein, and shares the

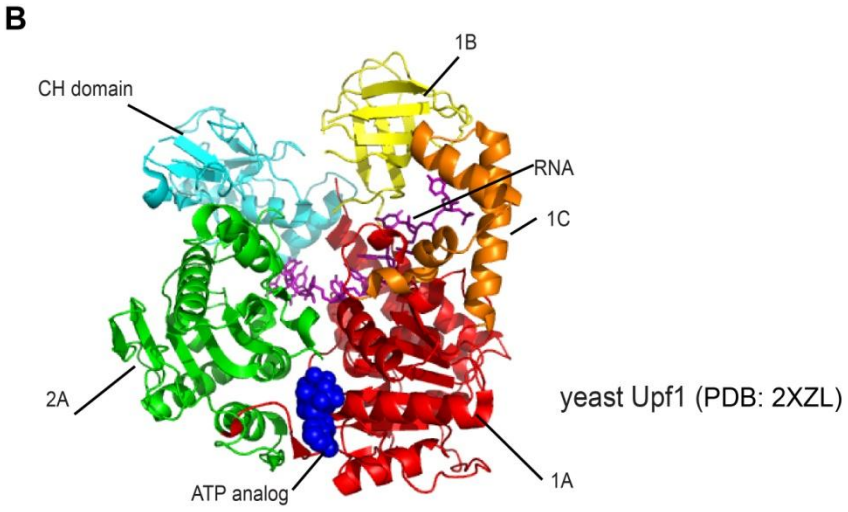
FIGURE 1.2 Model of the Sen1 termination complex adapted and modified from D. Brow (Brow 2011). The helicase Sen1 and two RNA-binding proteins, Nrd1 and Nab3, form the core complex, whereby Sen1 interacts with Nab3, and Nrd1 interacts with Nab3. Sen1 and Nrd1 both interact with the C-terminal domain of Pol II.


Rpb1-CTD
RNA
5'
Nrd1
Nab3
Sen1
Helicase
DNA
RNA Pol II
(Rpb1-12)

highest similarity with another SF1 helicase Upf1, allowing the crystal structure of the Upf1 helicase domain to be used as a model for study of Sen1 helicase before the crystal structure of Sen1 is solved (Figures 1.3A and B). Like other helicases, Sen1 has all the structurally and functionally conserved helicase motifs in its two tandem RecA domains. Residues in Motif I, II and VI are located on the cleft of the two RecA domains and responsible for ATP binding and hydrolysis. Motif 0 (Motif Q) and Motif IIIa are responsible for binding with the adenine base of the ATP molecule. Motif III and V function in the coordination between NTP and nucleic acid binding sites. The nucleic acid contacting surface is composed of Motif Ia-Ic, IV and V (FAIRMAN-WILLIAMS *et al.* 2010). In addition, Upf1 and Sen1 both have two large insertion domains, 1B and 1C, in the first RecA domain. The insertion domains are involved in nucleic acid binding in Upf1 (CHAKRABARTI *et al.* 2011), but their sequences between Upf1 and Sen1 are not conserved, and the function of Sen1's 1B and 1C domains is unknown. Another difference between the two helicase domains is that Sen1 does not have a cysteine and histidine enriched CH domain like Upf1. The CH domain is located right upstream of the core helicase domain and involved in interaction between Upf1 and its co-factors (CHAKRABARTI *et al.* 2011). Current evidences support the model that *S. cerevisiae* Sen1 utilizes its helicase activity to terminate transcription. An ortholog of Sen1 in the fission yeast *S. pombe* was found earlier to have ATP-dependent, 5' to 3' DNA and RNA unwinding activities *in vitro* (KIM et al. 1999). Using a simplified transcription termination system reconstituted *in vitro*, Porrua and Libri showed that native Sen1 purified from *S. cerevisiae* can elicit dissociation of paused Pol II from DNA *in vitro*, and its RNA-dependent ATPase activity is important for this function (PORRUA and LIBRI 2013). Recently, our lab proved that the recombinant *S. cerevisiae* Sen1 helicase domain has 5' to 3' helicase activity, as it exhibited ATP-independent single-stranded (ss) DNA

FIGURE 1.3 The helicase domains of Sen1, Upf1 and Senataxin are similar. A) Comparison of *S. cerevisiae* Sen1, *S. cerevisiae* Upf1 and human Senataxin. The conserved helicase domains and their subdomains are highlighted by colored boxes. Numbers indicate amino acid residues. The location of missense disease mutations in Sentaxin are shown by colored dots. B) Crystal structure of the *S. cerevisiae* Upf1 helicase domain in a complex with the non-hydrolyzable ATP analog, ADP:AlF_4^-, and RNA (PDB 2XZL, visualized by PyMOL, DeLano Scientific LLC) (CHAKRABARTI *et al.* 2011). The subdomains, ATP analog and RNA are indicated.


A
yeast Sen1
Q I Ia Ib Ic II III IIIa IV V VI
1
1146
1869
2231
1B 1A 1C 2A
ATPase-helicase domain
yeast Upf1
225
850
1
971
CH domain
human Senataxin
1
1736
2464
2677
(Disease substitutions in human Senataxin: AOA2 , ALS4)


B
1B
CH domain
RNA
1C
2A
yeast Upf1 (PDB: 2XZL)
ATP analog
1A

and RNA binding ability, and translocase activity on both in a 5' to 3' direction in the presence of ATP (S. Martin-Tumasz and D. Brow, unpublished data).

The importance of the Sen1 helicase domain for transcription termination is also revealed by the studies of helicase domain mutations *in vivo*. Before my study, there were only two missense mutations reported in the helicase domain of *S. cerevisiae* Sen1. The two hypomorphic alleles, named *sen1-1* and *nrd2-1*, have G1747D and E1597K substitutions, respectively (DEMARINI *et al.* 1992; STEINMETZ and BROW 1996). *sen1-1* was found from a genetic screen for temperature-sensitive mutants that are deficient in tRNA-splicing endonuclease activity *in vitro*, but later study suggests this defect could be caused by an indirect effect. *nrd2-1* was isolated from a genetic selection for read-through of a fortuitous Pol II-dependent terminator in the antisense strand of the U6 gene. Nevertheless, both mutations exhibit terminator read-through of a subset of Pol II-transcribed genes *in vivo*, indicating that the function of Sen1 helicase domain is required for efficient transcription termination by Pol II (STEINMETZ *et al.* 2001; STEINMETZ *et al.* 2006a; KUEHNER and BROW 2008; MISCHO *et al.* 2011; HAZELBAKER *et al.* 2013).

The N- and C-terminal regions outside the helicase domain of Sen1 are less conserved but still important for the functions of Sen1. The N-terminal 975 amino acids are dispensable for viability, however, there are multiple protein-protein interaction domains within the N-terminal region that are important for the functional specificity, including interactions with Rnt1 and Rpb1 (URSIC *et al.* 2004). The endoribonuclease Rnt1 functions in concert with Sen1 in 3' end formation of some sn/snoRNAs, and it may also mediate a fail-safe process for termination failure of some mRNA genes (CHANFREAU *et al.* 1998; URSIC *et al.* 2004; GHAZAL *et al.* 2009; RONDON *et al.* 2009). Sen1 interacts with the Pol II Rpb1 C-terminal domain (CTD), specifically

with the Serine2-phosphorylated (Ser2-P) CTD. The mutation in *SEN1* that disrupts this interaction caused reduced occupancy of the mutant Sen1 across the entire length of sn/snoRNA genes, as well as early and late in transcription of most mRNA genes (CHINCHILLA *et al.* 2012). Thus, this interaction could be important for Sen1 association with its target genes, but its exact role is unclear. In the C-terminal end of Sen1, there are a bipartite-like nuclear localization sequence and the Nab3 binding region that are important for its function in transcription termination (DEMARINI *et al.* 1992; URSIC *et al.* 1995; NEDEA *et al.* 2008). The yeast protein phosphatase Glc7 also directly interacts with the C-terminal end of Sen1 and is involved in snoRNA termination. The functional connection between Glc7 and snoRNA termination may involve dephosphorylation of Sen1 by Glc7 as shown *in vitro* (NEDEA *et al.* 2008).

Sen1 also utilizes its helicase activity to resolve co-transcriptional R loops. During transcription, R loops could arise when the newly-synthesized RNA strand threads back to the template DNA strand and forms a stable RNA:DNA duplex, leaving the non-template DNA single-stranded (ROY *et al.* 2010; SKOURTI-STATHAKI and PROUDFOOT 2014). The *sen1-1* mutant causes accumulation of R loops in a transcription-dependent manner, which induces DNA damage foci in cells and contributes to genome instability, indicating Sen1 functions to protect the heavily transcribed genes from the DNA damage associated with R loops (MISCHO *et al.* 2011; STIRLING *et al.* 2012). As discussed in Section 1.1.2, the function of resolving R loops by Sen1 also appears to be important for efficient transcription termination of some genes.

The RNA binding proteins Nrd1 and Nab3

The heterogeneous nuclear ribonucleoprotein (hnRNP)-like proteins Nrd1 and Nab3 are both essential for yeast cell viability, and each contains a conserved RNA recognition motif (RRM) (WILSON *et al.* 1994; STEINMETZ and BROW 1996; CARROLL *et al.* 2004). Terminator

read-through mutations have been found in both Nrd1 and Nab3, and biochemical experiments indicate they directly control transcription termination (STEINMETZ and BROW 1996; STEINMETZ and BROW 1998; STEINMETZ *et al.* 2001; KUEHNER and BROW 2008; LOYA *et al.* 2012). Nuclear depletion of Nrd1 resulted in genome-wide accumulation of transcripts from defective termination (SCHULZ *et al.* 2013; SCHAUGHENCY *et al.* 2014). In the prevailing model of the Sen1 termination complex, Nrd1 and Nab3 directly interact with each other to form a heterodimer, and further by the interaction between Sen1 and Nab3, the three components form a complex in yeast cells (Figure 1.2) (CONRAD *et al.* 2000; VASILJEVA and BURATOWSKI 2006; NEDEA *et al.* 2008; CHINCHILLA *et al.* 2012). The heterodimer formed by Nrd1 and Nab3 is postulated to reach higher order oligomerization using the C-terminal low complexity, self-assembly domain of Nab3, which is important for efficient termination at some Sen1 termination sites, possibly by increasing RNA-binding affinity (LOYA *et al.* 2013a; LOYA *et al.* 2013b).

Nrd1 and Nab3 appear to coordinate functions of other components involved in Sen1-dependent termination. By the interaction between Nrd1 and the Ser5-P CTD, the heterodimer of Nrd1 and Nab3 can bind to the elongation complex in early transcription, and recruit Sen1 by the Sen1 and Nab3 interaction in the core termination complex to elicit termination (see Section 1.2.3) (STEINMETZ and BROW 1998; CONRAD *et al.* 2000; NEDEA *et al.* 2008; VASILJEVA *et al.* 2008). Tudek *et al.* found that, by interacting with Trf4 of the TRAMP complex, Nrd1 may have a post-termination role of recruiting the TRAMP complex to execute 3' end processing (TUDEK *et al.* 2014). They found that this interaction is in a mutually exclusive manner with the Nrd1 and Pol II CTD interaction, and proposed that the Nrd1 and Nab3 heterodimer coordinates transcription termination and RNA degradation thereafter by switching between two alternative partners, which are the Pol II CTD for termination and the TRAMP complex for degradation.

The C-terminal domain of the Pol II Rpb1 subunit

The Pol II CTD can regulate Sen1-dependent termination by its phosphorylation states. The 26 (in yeast) or 52 (in human) heptad repeats with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ of the CTD provide the binding scaffold for transcription factors, and the phosphorylation states of the CTD during the transcription cycle determine its binding specificity with different factors (CORDEN 1990; WEST and CORDEN 1995). In the core Sen1 termination complex, Nrd1 has a CTD interaction domain (CID) preferentially bound to the Ser5-P CTD, which is enriched in promoter-proximal Pol II, and this interaction has been shown to be important for Sen1-dependent termination (STEINMETZ and BROW 1998; CONRAD *et al.* 2000; GUDIPATI *et al.* 2008; VASILJEVA *et al.* 2008). These evidences indicate that Nrd1 acts as a tether of the Sen1 termination complex to Pol II during early elongation, which allows Sen1-dependent termination to function on short non-coding RNA genes. The CTD phosphatases Ssu72 and Rtr1 remove the phosphate from the Ser5-P CTD in preparation for Ser2-P enrichment (KRISHNAMURTHY *et al.* 2004; MOSLEY *et al.* 2009). Ser2-P antagonizes Sen1-dependent termination and promotes poly(A)-dependent termination (GUDIPATI *et al.* 2008). This transition from the Ser5-P enriched CTD to the Ser2-P enrichment occurs when the nascent transcript is about ~450 nucleotides long, which defines a distance from the transcription start site for Sen1-dependent termination (ARNDT and REINES 2015).

Sen1-dependent termination read-through mutations were found in Ssu72, indicating that, not only the interaction between the Ser5-P CTD and Nrd1, but also the phosphorylation state transition of the CTD is critical for Sen1-dependent termination (STEINMETZ and BROW 2003). Perhaps the Ser5-P enriched CTD recruits Sen1 to prepare for termination by its interaction with the core termination complex. Then the transition in which Ser5-P de-phosphorylation and Ser2

phosphorylation occur could change the relative rates between the elongating Pol II and the Sen1 helicase, therefore allowing Sen1 to catch up with Pol II and elicit termination. In concert with this hypothesis, mutations that change the elongating rate of Pol II were shown to alter the location of Sen1 termination site depending on the relationship between the Pol II elongation rate and Sen1 translocation rate (HAZELBAKER *et al.* 2013).

The CTD residues other than Ser2 and Ser5 could also be involved in Sen1-dependent termination. Phosphorylation of Tyr1 impairs Nrd1 association with the CTD, therefore may have a role of anti-termination (MAYER *et al.* 2012). The Ser7-P CTD colocalizes with Nrd1 on transcription units that encode CUTs, SUTs and intronic snoRNAs, so this modification may functionally overlap with Nrd1 (KIM *et al.* 2010). In addition, the peptide bond orientation preceding the prolines in the CTD can be in either a *cis* or *trans* configuration. The conformation of the CTD is important for its interaction with other factors, for example, the interaction between Nrd1 and Ser5-P prefers the *cis* conformation of the Ser5-P-Pro6 peptidyl prolyl bond (KUBICEK *et al.* 2012). Ess1 is a prolyl isomerase that catalyzes the *cis*/*trans* conversion of Ser5-P/Pro6 and Ser2-P/Pro3 bonds of the CTD, and appears to participate in Sen1-dependent termination by coordinating the recruitment of Nrd1 and other termination factors (SINGH *et al.* 2009). Collectively, the CTD has a complex control on Sen1-dependent termination.

1.3. The human ortholog of yeast Sen1, Senataxin

Mutations in *SETX* co-segregate with neurodegenerative diseases

The human ortholog of yeast Sen1, called Senataxin, is encoded by the gene *SETX*. Mutations in *SETX* co-segregate with two rare familial, progressive neurodegenerative disorders with juvenile onset: autosomal recessive ataxia oculomotor apraxia type 2 (AOA2) and

autosomal dominant amyotrophic lateral sclerosis type 4 (ALS4) (Figure 1.3A) (CHEN *et al.* 2004; MOREIRA *et al.* 2004; LEMMENS *et al.* 2010). AOA2 is characterized by severe cerebellar atrophy, gait ataxia, peripheral neuropathy, elevated serum alpha-fetoprotein level, and about 50% of cases show oculomotor apraxia (FOGEL *et al.* 2014). The mutations that give rise to AOA2 are diverse, including missense, nonsense, insertion/deletion, splice site and frameshift mutations. ALS4 is more rare than AOA2 and is associated with severe muscle weakness and pyramidal signs, but unlike classical ALS, it does not affect respiratory muscles or lifespan of the patients (ARNING *et al.* 2013). ALS4 mutations identified so far are exclusively missense (FOGEL *et al.* 2014). The mechanisms by which the *SETX* mutations cause AOA2 and ALS4 have not been understood. Although a model whereby loss-of-function results in the recessive AOA2 and gain-of-function causes the dominant ALS4 can be proposed based on the natures of the disease-causing mutations, our study of the AOA2 mutations showed dominant effects caused by some of the mutations, suggesting a more complex situation at least for AOA2 (CHEN *et al.* 2014). About half of the missense disease mutations are found to locate in the conserved helicase domain (Figure 1.3A), suggesting the diseases are associated with dysfunction of the helicase domain.

One possible cause of AOA2 and ALS4 is altered gene expression by the disease mutations in Senataxin. Senataxin has been implicated in Pol II transcription termination on protein-coding genes (SKOURTI-STATHAKI *et al.* 2011). Gene expression profiling showed that AOA2 mutations in *SETX* cause changes in gene expression in patient fibroblasts, including genes involved in neurogenesis and neuronal functions (FOGEL *et al.* 2014). Thus, the two neurodegenerative diseases could result from defects in expression of genes essential for long-term neuron survival. DNA damage triggered by R loops may also contribute to the pathogenesis

process of AOA2 and ALS4, as suggested by the observation that AOA2 cells showed increased sensitivity to DNA damage reagents and defective double-strand break (DSB) repair, which can be corrected by full-length *SETX* cDNA (SURAWEERA *et al.* 2007).

Implication of Senataxin in cancer

Recently, the role of Senataxin in resolving R loops has been implicated in cancer development. Senataxin was identified as an interactor of the breast cancer susceptibility gene 1 (BRCA1) from two independent, systematic screens (HILL *et al.* 2014). Further experiments showed that the complex of Senataxin-BRCA1 is recruited to the transcription termination pause sites of some highly transcribed genes and suppresses R-loop-associated DNA damage there. Furthermore, insertion/deletion somatic mutations were found near some of these transcription termination pause sites in breast tumor tissues carrying inherited BRCA1 mutations (HATCHI *et al.* 2015). This discovery highlighted the potential role of the Senataxin-BRCA1 complex in preventing cancer by resolving R loops and protecting the genome from instability, but current evidence that ALS4/AOA2 patients are more susceptible to cancer is indefinite.

1.4. Dissertation organization

In this dissertation, I discuss my studies of the mechanism and targets of Sen1 helicase in termination of non-coding RNA synthesis in *S. cerevisiae*. In **Chapter 2**, I describe the materials and methods of the experiments conducted in this dissertation. In **Chapter 3**, I discuss my investigation of identifying the minimal essential region of Sen1 and the effects caused by AOA2 disease mutations on Sen1-dependent termination. Using a yeast genetic system, I identified the helicase domain plus a flanking nuclear localization sequence as the minimal essential region of Sen1. Guided by the structure model of the yeast Upf1 helicase domain, I found the molecular defect underlying a termination read-through mutation *sen1-E1597K* is disruption of an

intradomain ionic bond of the helicase domain. Since human Senataxin cannot replace Sen1 in *S. cerevisiae*, I tested 13 AOA2-associated *SETX* mutations in the context of yeast *SEN1*, and showed most of them cause growth and termination defects. In **Chapter 4**, I discuss my investigation of the targets of the Sen1-dependent transcription termination pathway. Hypomorphic mutations in proteins involved in Sen1-dependent termination, including Sen1, Nrd1, Nab3, Ssu72, Rpb11 and Hrp1, were used to characterize the influences of these six proteins on Sen1-dependent termination across the genome. I found the genome-wide effects caused by the mutations are surprisingly diverse. The substitution in Sen1 induces the broadest and strongest termination defects, while the other substitutions only affect efficient transcription termination of different subsets of Sen1's targets. I confirmed that the six proteins participate in transcription termination of sn/snoRNA genes, but to different degrees, and showed that the RNA-binding abilities of Nrd1 and Nab3 are dispensable for termination of some snoRNA genes, such as *SNR39B*. I also revealed that Sen1 may regulate meiotic gene expression by repressing their transcription in vegetative cells. In **Chapter 5**, I discuss the implications and future directions based on the results presented in Chapters 3 and 4. In **Appendix 1**, I discuss my genetic selections for Sen1 terminator read-through mutations in the helicase domain of Sen1. In **Appendix 2**, I describe my experiments of observing the GFP-labeled wild-type Sen1 in live yeast cells using time-lapse fluorescence microscopy. In **Appendix 3**, I describe my experiments of verifying a previously isolated terminator read-through mutant with a premature stop codon mutation in *NAB3*, and creating a plasmid-based yeast system to characterize this mutant. The work presented in Chapter 3 has been published in Chen *et al.* Genetics. 2014.

CHAPTER 2: MATERIALS AND METHODS

2.1 Plasmid construction

The pRS425-SEN1(PstI) plasmid was made by Tom Gisel and Dave Brow. The *SEN1* gene, including 434 bp upstream of the start codon and 212 bp downstream of the stop codon, was amplified from genomic DNA of the strain 46**α** in two halves joined at an internal *Pst*I site created via a silent mutation in codons 964-966. The PCR products were inserted into SalI/SpeI-cut pRS425 vector to make the first *SEN1* shuffle-plasmid construct (CHEN *et al.* 2014). With help from Elaine Brow, Christine Treba and Ulrika Müller, I subcloned the 7,342 base pair *SEN1* gene into pRS313, pRS315 and pRS316, and created point mutations in *SEN1* in pRS313 by the QuikChange procedure (Stratagene).

Deletions of *SEN1* were created using pRS313-SEN1(PstI). The C-terminal sequences were deleted by the QuikChange procedure, leaving the stop codon intact. The initial N-terminal deletion was made by deleting codons 4-964 of the *SEN1* ORF and adding a silent mutation to codon 3 to recreate the internal *Pst*I site. In this construct, codons 3, 965 and 966 have the sequence T<u>CT-GCA-G</u>AA, with the *Pst*I site underlined. In subsequent deletions, codons 1-3 and 965 were retained, and codon 966 was replaced with the indicated codon (e.g., 1004, 1089, etc.), which always begins with a G to retain the *Pst*I site. The presence of this *Pst*I site as well as one in the pRS313 vector sequence allowed construction of each plasmid by ligation of two PCR fragments. All deletion plasmids were sequenced across the deletion junction to confirm their identity. C-terminal deletions were added to N-terminal deletion constructs using the *Nco*I site at codons 1890-1892 and the downstream *Spe*I site.

Sen1-GFP fusion constructs were made from pRS315-SEN1(PstI). Using the QuikChange procedure, an internal *Avr*II site was created by silent mutation of codon 2230 to generate pRS315-SEN1(AvrII). Codons 2 through 238 of the GFP(S65T) variant were amplified

from genomic DNA of the yeast strain KES000 (made by Kaitlin Sundling) containing an *IMD2-GFP* gene fusion (derived from Life Technologies #95700). *AvrII* sites were created at both ends of the GFP(S65T) coding sequence to allow insertion into pRS315-SEN1(AvrII), generating pRS315-SEN1-GFP. N-terminal deletions of *SEN1* in pRS313 were subcloned into pRS315-SEN1-GFP. To make GFP fusions of C-terminally deleted *SEN1*, sequences upstream of the *AvrII* site were deleted from pRS315-SEN1(AvrII) by the QuikChange procedure and the GFP(S65T) fragment was inserted into the *AvrII* site afterwards. The only exception is that the C-terminal deletion after codon 1929 was made by inverse PCR from pRS315-SEN1-GFP and blunt-end ligation, removing the upstream *Avr*II site in the process.

The Senataxin/Sen1 chimeric constructs were made from pRS313-SEN1(PstI). For the entire *SEN1-SETX* ORF swap construct, *SEN1* codons 4-2232(stop) were deleted and replaced with a GGA (Gly) codon to create a *BspE*I site by inverse PCR, digestion, and self-ligation to make the plasmid pRS313-SEN1ORFdel(BspEI). *SETX* codons 3-2678(stop) were PCR-amplified from pSETX (SURAWEERA *et al.* 2007) with a *BspE*I site at both ends. The PCR product was digested with *BspE*I and ligated into the pRS313-SEN1ORFdel(BspEI) plasmid to generate pRS313-SEN1/SETXORF. For the *SEN1-SETX* helicase domain swap construct, unique *Pml*I and *Nco*I sites were used to excise *SEN1* codons 1175-1890 and replace them with *SETX* codons 1769-2484 to generate pRS313-SEN1/SETXhelicase. For the *SEN1-SETX* Rec1A domain swap construct, unique *Pml*I and *BspE*I sites were used to excise *SEN1* codons 1175-1474 and replace them with *SETX* codons 1769-2069. Two constructs were made for the *SEN1-SETX* Rec2A domain swap, unique *BspE*I and *Nco*I sites, or *Bsg*I and *Nco*I sites were used to excise *SEN1* codons 1478-1889 or 1635-1889, and replaced them with *SETX* codons 2073-2483 or 2226-2483, respectively. For the *SEN1-SETX* 1B domain swap construct, *SEN1* codons 1174-

1285 were deleted and replaced with an *Sph*I digestion site. *SETX* codons 1770-1889 were PCR amplified with an *Sph*I site at both ends. The PCR product was digested with *Sph*I and inserted into the *Sph*I-digested *SEN1* 1B domain deletion construct. Other domain swap constructs were created by the same strategy. For the *SEN1-SETX* 1C domain swap construct, *SEN1* codons 1451-1565 were replaced with *SETX* codons 2057-2150 using an *Sph*I digestion site. For the *SEN1-SETX* 1B-1C domain swap construct, *SEN1* codons 1174-1285 and 1455-1562 were replaced with *SETX* codons 1770-1889 and 2062-2145 using an *Sph*I and a *BsrG*I digestion site respectively.

For the *CUP1* termination efficiency reporters, pGAC24-CYC1 (STEINMETZ and BROW 2003) and pRS316-NRD1-CUP1 (KUEHNER and BROW 2008) were made previously. The pGAC24-SNR47 reporter plasmid was generated by amplifying the *SNR47* terminator (CARROLL *et al.* 2004) (positions +100 to +215, relative to +1 transcription start site) from DAB206 (see Section 2.2) genomic DNA with an *Xho*I site introduced at each end. The PCR product was digested with *Xho*I and ligated to XhoI–cut pGAC24 (LESSER and GUTHRIE 1993).

2.2 Yeast Strains

Plasmids were transformed into *S. cerevisiae* by the lithium acetate procedure (SCHIESTL and GIETZ 1989). The wild-type plasmid-shuffle yeast strain DAB206 (*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS316-*SEN1*]) is derived from 46α (*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2*) (LESSER and GUTHRIE 1993) and made by Ulrika Müller (CHEN *et al.* 2014). pRS316-SEN1 was replaced with pRS313-borne *SEN1* alleles by standard plasmid shuffle techniques (BOEKE *et al.* 1987). Loss of pRS316-*SEN1* was selected for on synthetic complete medium containing 0.75 mg/mL of 5-fluoroorotic acid (5-FOA).

XCY361 (*MATα his3-Δ1 leu2-Δ0 lys2-Δ0 ura3-Δ0 NHP6A-TagRFP-T:nat*) was derived

from BY4742 (Research Genetics, Inc.) by integrating the *NHP6A* gene fused with TagRFP-T followed by the nourseothricin resistance (*nat*) gene into the endogenous *NHP6A* locus. The NHP6A-TagRFP-T/*nat* DNA fragment used for chromosomal integration was generated by Kaitlin Sundling (CHEN *et al.* 2014). Integration at the *NHP6A* locus was confirmed by PCR of genomic DNA with flanking oligos and sequencing.

For the microarray analysis, the *sen1-E1597K* (*nrd2-1*) (*MAT***a** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1-E1597K*), *nrd1-V368G* (*nrd1-5*) (*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nrd1-V368G*), *nab3-F371L/P374T* (*nab3-11*) (*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS313-*nab3-11(F371L/P374T)*]), *ssu72-G33A* (*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 ssu72-G33A*), *rpb11-E108G* (*MAT***a** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rpb11-E108G*) and *hrp1-L205S* (*hrp1-5*) (*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hrp1::HIS3* [pRS315-*hrp1-5(L205S)*]) strains and their wild-type parent strains 46a and 46α were described previously (STEINMETZ and BROW 1996; STEINMETZ and BROW 2003; STEINMETZ *et al.* 2006a; KUEHNER and BROW 2008). 46**a** is isogenic to 46**α** but has the opposite mating type.

The Nrd1/Nab3 double mutant (*nrd1-5/nab3-11*) (*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nrd1-V368G nab3::KanMX* [pRS313-*nab3-11(F371L/P374T)*]), named XCY377, was created by homologous recombination integrating *nab3::KanMX* amplified from YSC1021-669432 (Open Biosystems) into the *nrd1-V368G* (*nrd1-5*) genome transformed with pRS316-NAB3 (named XCY373), and then standard plasmid shuffle techniques. Integration at the *NAB3* locus was confirmed by PCR of genomic DNA with flanking oligos and sequencing.

2.3 Protein Analysis

Rabbit antiserum against amino acids 1095-1876 of Sen1 was prepared by Steve Martin-Tumasz (Chen et al. 2014). IgG was purified from antiserum using the Melon^{TM} Gel IgG Spin

Purification Kit (Thermo Scientific). For immunoblots of Sen1, 10 OD units of yeast cells growing logarithmically (OD_{600} 0.5-1.0) at 30 ℃ in YEPD (1% yeast extract/2% peptone/2% dextrose) were pelleted by centrifugation, lysed by vortexing with glass beads (425-600 μm diameter; Sigma-Aldrich) in 100 μl phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) with Protease Inhibitor Cocktail P2714 (Sigma-Aldrich), 100 μl of 2× SDS PAGE loading buffer (100 mM Tris-Cl pH 6.8, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM β-mercaptoethanal) was added, and the mixture was heated at 100 ℃ for 5 min. Samples were electrophoresed using 4-15% Mini-PROTEAN TGX Precast Gels (Bio-Rad) at 160 V for 50 min, and electroblotted onto Hybond-C extra nitrocellulose membranes (Amersham Biosciences) at 6 V/cm, 4 ℃ for 12 hours in 20% (v/v) methanol in Towbin buffer (25 mM Tris base, 192 mM glycine, pH 8.3). Blots were blocked with 5% dried milk in PBS buffer with 0.1% Tween-20 at 23 ℃ for 1 hour, incubated with purified Sen1 IgG (1:2,000 dilution) for 1 hour, and then Horse Radish Peroxidase-conjugated Goat anti-Rabbit IgG (Thermo Scientific) (1:50,000 dilution) for 1 hour. Blots were developed with Immobilon™ Western Chemiluminescent HRP Substrate (Millipore) and exposed to HyBlot CL Autoradiography Film (Denville Scientific). Immunoblots of Nrd1 were performed by the same protocol except the Nrd1 polyclonal antiserum (STEINMETZ and BROW 1998) was diluted 1:1,000.

2.4 Microscopy

Culture of yeast cells, microscopic imaging and image processing for the studies described in Chapter 3

XCY361 cells transformed with GFP-tagged *SEN1* constructs in pRS315 were grown in synthetic complete (SC) medium lacking leucine at 30 ℃ to OD_{600} = 0.5, then diluted to OD_{600} =

0.2. Two hundred microliters of culture was placed on the center of a glass coverslip pre-coated with 1 mg/mL Concanavalin A, Type IV (Sigma-Aldrich) solution and viewed with a Nikon Eclipse Ti fluorescent microscope with a Nikon DS-Qi1Mc digital camera and NIS-elements AR 4.12.01 software. Filter sets with excitation/emission wavelengths of 465-495 nm/515-555 nm and 530-560 nm/573-648 nm were used for the GFP images and RFP images, respectively. Image files were processed and analyzed by ImageJ (NIH) following the ImageJ User Guide (http://rsbweb.nih.gov/ij/docs/guide/146.html). For background subtraction, the command "Process>Subtract background>Create Background" was used to create the background, and the command "Process>Image Calculator" was used to subtract background from the original image. For contrast enhancement, the command "Image>Adjust>Brightness/Contrast" was used. In most cases, appropriate contrast was achieved by a single click on "Auto".

Culture of yeast cells, ONIX microfluidic system, microscopic imaging and image processing for the studies described in Appendix 2

XCY361 cells transformed with GFP-tagged *SEN1* constructs in pRS315 were grown in SC medium lacking leucine at 30 ℃ to saturation. A portion of the culture was diluted to OD_{600} = 0.2 in SC medium lacking leucine and grown at 30 ℃ to OD of 0.35-0.6 before loading into the ONIX system.

The ONIX-262 microfluidic system with Y04C microfluidic plates for haploid yeast cells (CellAsic) was used to trap cells for the time-lapse imaging experiment. This system consists of a computer-operated control system, a manifold that connects a vacuum pump and seals to the top of the plate, and a disposable microfluidic plate (see the product instruction online: http://www.emdmillipore.com/US/en/product/CellASIC-ONIX-Complete-Perfusion-System,MM_NF-EV262?bd=1#overview). The manifold can adjust the pressure in designated

wells of the microfluidic plate, so that cells in the wells can be loaded into the culture chambers in the center of the plate for imaging, and continuous flow of media is maintained over the entire experiment. Inside the culture chamber, yeast cells are trapped by an elastic ceiling, which immobilizes the cells and maintains them in a uniform focal plane for imaging throughout the experiment. Cells were loaded until there were enough cells in the imaging field, and washed with 5 psi SC medium lacking leucine for 5 minutes to remove the untrapped cells. A flow pressure of 1 psi was applied for the experiment after washing. The microfluidic experiment was conducted at room temperature.

A Zeiss Axiovert 200M inverted microscope with AxioCam HR CCD camera from David Wassarman's lab, University of Wisconsin-Madison, was using for imaging cells. I used a 40X oil immersion Plan Neofluar objective with a numerical apterture of 1.3. A halogen light source and phase 2 illumination setting were used for the phase contrast images. An HBO fluorescence light source was used for fluorescence imaging. Filterset 38 (eGFP) from Zeiss with excitation/emission wavelengths of 450-490 nm/500-550 nm was used for the GFP channel. For each time course experiment, all the images were taken in the same imaging field. A phase contrast image and a GFP image were taken every time point. The exposure time for GFP imaging is 1.5 seconds. Phase contrast images were used to generate the cell outline, and GFP images were taken for Sen1-GFP.

Image files were processed and analyzed by ImageJ (NIH) following the ImageJ User Guide (http://rsbweb.nih.gov/ij/docs/guide/146.html). For GFP image background subtraction, the command "Process>Subtract background>Create Background" was used to create the background, and the command "Process>Image Calculator" was used to subtract background from the original image. For phase contrast image processing, the command

"Image>Adjust>Brightness/Contrast>Set" was used, and the display value was set to the range of -600 to 1200. The processed GFP and phase contrast images of the same time point were then merged using the command “Image>Color>Merge channels”. The resulting merged image was stacked by “Stack to RGB”, and a satisfying contrast could be reached by setting the display value to the range of 50 to 200. The images of time series can be imported by ImageJ using “Image Sequence” and saved as a movie file.

2.5 Total RNA preparation

These RNA samples were used for the studies described in Chapter 3 and Appendix 1.

Yeast cells were grown in YEPD at 30 ℃ to OD_{600} = 0.8-1.0, quickly shifted to 37 ℃ by mixing with the same volume of 44 ℃ YEPD, continued growing for 1 hour and pelleted by centrifugation. Total cellular RNA was prepared by the glass bead-guanidinium isothiocyanate-hot phenol extraction method (WISE 1991).

These RNA samples were used for the studies described in Chapter 4 except Figure 4.7.

RNA was prepared by the Auble lab, University of Virginia Health System, Charlottesville (UVA). Strains were grown in YEPD at 30 ℃, then shifted to 35 ℃ for 45 minutes and pelleted by centrifugation, except the Rpb11 mutant, which was grown continuously at 30 ℃. Total cellular RNA was prepared by the method of Schmitt *et al.* (SCHMITT *et al.* 1990).

These RNA samples were used for the studies described in Figure 4.7.

Yeast cells were grown in YEPD at 23 ℃ to OD_{600} = 0.8-1.0, quickly shifted to 37 ℃ by mixing with the same volume of 51 ℃ YEPD, continued growing for 1 hour and pelleted by centrifugation. Total cellular RNA was prepared by a method adapted from Schmitt *et al.* (SCHMITT *et al.* 1990) with the following modifications: Saturated phenol (pH 4.3 $\pm$ 0.2) (Fisher) was used for the hot phenol extraction step. The extraction reaction was heated to 65 ℃ in water

bath for 6 minutes, and vortexed every 2 minute for 30 seconds.

2.6 Northern blot

Random-primed radioactive probe preparation for the studies described in Chapter 3 and Appendix 1

The random-primed radioactive probes for *NRD1* and *SNR47* were generated using gel-purified PCR products as templates. The probe labeling protocol was adapted from the Prime-a-Gene® Labeling System (Promega) using [α-^{32}P]dTTP (PerkinElmer). After labeling, unincorporated nucleotides were removed by a G-50 Sephadex spin column (GE Healthcare).

Radioactive oligonucleotide probe preparation for the studies described in Chapter 4

The oligonucleotide probes for *SNR39B, SNR47, FLO1, CWP1, CWP2* and *NDJ1* were made by Integrated DNA Technologies and are described in Table 2.1. USB OptiKinase™ (Affymetrix) was used to label the oligonucleotides with [γ-^{32}P]ATP (PerkinElmer) at 37 ℃ for 1 hour. After labeling, unincorporated nucleotides were removed by a G-50 Sephadex spin column (GE Healthcare).

For Northern blots, 15–20 μg RNA was loaded in each lane, run on a 0.8% or 1% agarose/0.8 M formaldehyde gel and blotted onto Zeta-probe membrane (Bio-Rad) by capillary transfer.

For hybridization using random-primed probes, the blot was first pre-hybridized in Ultrahyb solution (Ambion) at 42 ℃ for 1 hour, then hybridized in Ultrahyb solution with probe concentration about 10^6 cpm/mL at 42 ℃ for over 12 hours.

For hybridization using oligonucleotide probes, the blot was first pre-hybridized in Church buffer (0.2 mM Na_2HPO_4, 0.1 mM NaH_2PO_4, 1 mM EDTA, 7% SDS, 10 mg/mL BSA) at 60 ℃ for 1 hour, then hybridized in Church buffer with probe concentration about 10^6 cpm/mL at 60 ℃

for over 12 hours.

After hybridization, the blot was washed in saline-sodium citrate (SSC) wash buffers (2× SSC, 0.1% SDS, 5 min per wash, twice, at 23 ℃; 0.1× SSC, 0.1% SDS, 15 min per wash, twice, at 42 ℃), and visualized with a Phosphorimager (Molecular Dynamics).

2.7 Random mutagenesis selection for read-through mutations in the Sen1 helicase domain

Random mutagenesis using error-prone PCR by Taq DNA polymerase (NEB) was performed to introduce mutations into the regions of interest in Sen1. For each selection, the gapped pRS313-SEN1 plasmid indicated below, the mutagenic PCR product of *SEN1*, and the *CUP1* reporter construct were co-transformed into DAB206. His^+Leu^+ transformants were replica plated to medium containing 5-FOA to select for cells that had lost the URA3-marked wild-type *SEN1* allele, and surviving colonies were replica plated to medium containing copper to select for cells with substitutions that promote terminator read-through in the *CUP1* reporter constructs (0.2 mM copper sulfate for the first selection using pGAC24-CYC1, and 0.1 mM copper sulfate for the last two selections using pGAC24-SNR47). Cells were always incubated at 30 ℃ during the selections. Genomic DNA of the surviving colonies was extracted. I PCR-amplified the mutagenized region from the genomic DNA using PfuTurbo DNA polymerase (Strategene), and sequenced the PCR products to identify the potential read-through mutations.

For the random mutagenesis selection of the C-terminal half of the helicase domain, *SEN1* codons 1477 to 1890 were subjected to mutagenesis. A gapped plasmid was generated by digesting pRS313-SEN1(PstI) with *BspE*I/*Nco*I. Randomly-mutagenized *SEN1* fragments were generated using the oligo pairs SEN1-4406F/SEN1-5R and SEN1-3F/SEN1-5R (See Table 2.1) with wild-type *SEN1* as template. The *CUP1* reporter construct for this selection was pGAC24-CYC1.

For the random mutagenesis selection of the 1B domain, *SEN1* codons 1174-1285 were subjected to mutagenesis. The gapped plasmid was the *Sph*I-digested *SEN1* 1B domain deletion construct that was used for making the *SEN1-SETX* 1B domain swap. The randomly-mutagenized *SEN1* fragments were generated using the oligo pair SEN1-1B-upstream and SEN1-1B-downstream (See Table 2.1) with wild-type *SEN1* as template. The *CUP1* reporter construct for this selection was pGAC24-SNR47.

The random mutagenesis selection of the 1C domain was performed by the same method as the 1B domain. *SEN1* codons 1451-1565 were subjected to mutagenesis. The gapped plasmid was the *Sph*I-digested *SEN1* 1C domain deletion construct, and the oligo pair SEN1-1C-upstream and SEN1-1C-downstream (See Table 2.1) was used for error-prone PCR.

2.8 Total RNA genomic tiling array and data analysis

For the microarray experiment, total RNA extraction, cDNA synthesis, fragmentation, labeling and hybridization to GeneChip® *S. cerevisiae* Tiling 1.0R Arrays (Affymetrix, Inc., purchased by the Brow lab) were conducted by the Auble lab at UVA using the protocol reported previously (POOREY *et al.* 2010). Raw data were generated by the Microarray Core Facility at UVA. The RNA analyses of both 46**a** and 46**α** wild-type strains, and the *sen1-E1597K, nrd1-V368G*, *nab3-F371L/P374T*, *ssu72-G33A*, *hrp1-L205S* and *rpb11-E108G* mutants were performed using two independent biological replicates for each strain.

I used the genomic map (BPMAP) file in Affymetrix® Tiling Analysis Software (TAS) version 1.1 to analyze intensity data stored in the raw microarray CEL files and evaluate the signal for each interrogated genomic position. The BPMAP file contains information about every probe's x,y location on the chip and its location on a chromosome. Because the *S. cerevisiae* Tiling 1.0R Array is drawn from the reference genome sequence UCSC sacCer1 (released Oct.

2003), and there have been changes in the later versions, I used the UCSC sacCer3 sequence (released Apr. 2011) to update the BPMAP file so that every probe on the chip that has a unique perfect match in the yeast genome released most recently can be assigned to its correct chromosomal location. The BPMAP file was updated by the function "remap" of the Bioconductor software package "Starr".

The "absolute" RNA level of each sample was calculated following the TAS user guide available on the Affymetrix website. The intensities on the arrays of the two biological replicates were quantile normalized, and then the data were linearly scaled to a target median intensity of 100. The *S. cerevisiae* Tiling 1.0R Arrays has paired perfect match (PM) and mismatch (MM) probes, and MM probes serve as background. TAS applied the Wilcoxon Signed-Rank test to the normalized log2(max(PMi-MMi,1)) value whose genomic coordinate i fell within a designated 101-bp smoothing window to calculate the log transformed probability (-10log10(p-value)) that the RNA was detected above the noise level. The associated Hodges-Lehmann estimator was used to compute the signal intensity. If the estimate s of signal location is over 1, it was transformed to $\log_2(s)$ for the RNA level, otherwise, 0 was assigned for the RNA level.

Similarly, the differential RNA level was estimated by quantile normalizing replicate arrays of each strain and scaling the data to a target median intensity of 100 with mutant RNA samples as treatment and the wild-type RNA samples as control. TAS generated estimates of fold change by performing the Wilcoxon Rank-Sum test and calculating the associated Hodges-Lehmann estimator within a designated smoothing window. The difference between the treatment estimate s (s_t)A and the control estimate s (s_c), $s_t - s_c$, of signal location is interpreted as the log2 fold change between the treatment and control group signals. A 101-bp smoothing window was used for all the differential RNA level analyses, except that the discrete loop identification used a 500-

bp smoothing window.

The output "absolute" and differential RNA level data are in the readable format with intensity value for each interrogated genomic position. All the genome features used for later data visualization and computation in this study are downloaded from the *Saccharomyces* Genome Database (SGD). Because the *HIS3* gene in the *nab3-F371L/P374T* mutant and the *LEU2* gene in the *hrp1-L205S* mutant are carried in the plasmids, to avoid the interference from plasmid amplification, I removed the data points that fall into the *HIS3* and *LEU2* regions (Chr15: 771,766-722,818 and Chr3: 91,185-92,702) from later analyses. This process generated RNA level values of 2,585,209 genomic locations across the whole genome for every experimental group. For data visualization, I used Integrated Genome Browser (IGB) (NICOL *et al.* 2009). I used MATLAB (MathWorks, Inc.) for large-scale data computation. Data points with values equal to zero suggest their PM signals are not above the noise, and therefore were precluded during calculation.

TABLE 2.1 Oligonucleotides used in this dissertation

Primer Name	Primer Sequence (5' - 3')
Mutagenesis	
SEN1 E1597K F	GAAGCATGTCAATGTACCAAACTTTCTTCAATTATTCCTCTC
SEN1 E1597K R	GAGAGGAATAATTGAAGAAAGTTTGGTACATTGACATGCTTC
SEN1 G1747D F	GATTTCACGGGAAAAATCGATATAATTTCACCTTATAGAGAG
SEN1 G1747D R	CTCTCTATAAGGTGAAATTATATCGATTTTTCCCGTGAAATC
W1166S-up	ACGCTTTACTACTTCTAGAATCTTCGCAAGGTCTATGTTCTTCAC
W1166S-down	GTGAAGAACATAGACCTTGCGAAGATTCTAGAAGTAGTAAAGGCT
Sen1 I1370R fwd	AGGATTGGTTACTTTTTGTCTACAAAGAATGC
Sen1 I1370R rev	ACCCAGTATAGTCTTTGTTTTACCCGTAC
I1371F-up	AACAAAGACTATACTGGGTATTTTTGGTTACTTTTTGTCTACAAAG
I1371F-down	CTTTGTAGACAAAAAGTAACCAAAAATACCCAGTATAGTCTTTGTT
SEN1-C1409Y-F	CTGTTAAAAAAGCAAAAAATTCTAATCTATGCCCCCAGTAATGCC
SEN1-C1409Y-R	GGCATTACTGGGGGCATAGATTAGAATTTTTTGCTTTTTTAACAG
SEN1-N1413S-F	CTAATCTGTGCCCCCAGTAGTGCCGCTGTGGACGAAATC
SEN1-N1413S-R	GATTTCGTCCACAGCGGCACTACTGGGGGCACAGATTAG
SEN1-L1569W-F	GTGACATCATATGTTCGACTTGGTCTGGTTCAGCCCATG
SEN1-L1569W-R	CATGGGCTGAACCAGACCAAGTCGAACATATGATGTCAC
SEN1-D1616V-F	CGTTGTATTATGGTTGGTGTTCCTAACCAACTACCAC
SEN1-D1616V-R	GTGGTAGTTGGTTAGGAACACCAACCATAATACAACG
SEN1 P1622L F	CCTAACCAACTACCACTAACTGTTCTTTCAGGTGCA
SEN1 P1622L R	TGCACCTGAAAGAACAGTTAGTGGTAGTTGGTTAGG
F1767L-up	GAGTTTGCCCGCTATTTAGGAGGCATGATCAACAA
F1767L-down	TTGTTGATCATGCCTCCTAAATAGCGGGCAAACTC
SEN1-T1779P-F	ATCAACAAATCAATCGATTTCAATCCAATCGACGGTTTCCAAG
SEN1-T1779P-R	CTTGGAAACCGTCGATTGGATTGAAATCGATTGATTTGTTGAT

K1788E forward	CGGTTTCCAAGGTCAAGAAGAAGAAATCATCTTGATATCGTGTG
K1788E reverse	CACACGATATCAAGATGATTTCTTCTTCTTGACCTTGGAAACCG
R1820Q forward	GTATGAATGTGGCTCTGACTCAAGCCAAGACAAGTATTTGG
R1820Q reverse	CCAAATACTTGTCTTGGCTTGAGTCAGAGCCACATTCATAC
SEN1 R1850H F	GATTTGATTGAAGATGCAAAAGATCATAGTTGTCTTGCATATGCTTGC
SEN1 R1850H R	GCAAGCATATGCAAGACAACTATGATCTTTTGCATCTTCAATCAAATC
Del1908-2231 up	GAGGTAAAATCCAATAAGGACACATGATCTTGAAGTAAAAGTGTTAGG
Del1908-2231 down	CCTAACACTTTTACTTCAAGATCATGTGTCCTTATTGGATTTTACCTC
Del1952-2231 up	GAATAACAAGAAGGCAGAATCGTGATCTTGAAGTAAAATTGTTAGG
Del1952-2231 down	CCTAACACTTTTACTTCAAGATCACGATTCTGCCTTCTTGTTAATC
Del1984-2231 up	CTTTTCCTGATGTGGACAGCTGATCTTGAAGTAAAAGTGTTAGG
Del1984-2231 down	CCTAACACTTTTACTTCAAGATCAGCTGTCCACATCAGGAAAAG
Del2097-2231 up	CGACGATCCACAAGTTTCGTGATCTTGAAGTAAAAGTGTTAGG
Del2097-2231 down	CCTAACACTTTTACTTCAAGATCACGAAACTTGTGGATCGTCG
Del 1859-2231 up	CTTGCATATGCTTGCTCGTGATCTTGAAGTAAAAGTGTTACC
Del 1859-2231 down	CCTAACACTTTTACTTCAAGATCACGAGCAAGCATATGCAAG
Sen1-1907-GFP-fwd	GAGGTAAAATCCAATAAGGACACACCTAGGAGTAAAGGAGAAGAAC
Sen1-1907-GFP-rev	GTTCTTCTCCTTTACTCCTAGGTGTGTCCTTATTGGATTTTACCTC
Sen1-1984-GFP-fwd	CTTTTCCTGATGTGGACAGCCCTAGGAGTAAAGGAGAAGAAC
Sen1-1984-GFP-rev	GTTCTTCTCCTTTACTCCTAGGGCTGTCCACATCAGGAAAAG
Sen1-1907-GFP-f-AvrII	GAGGTAAAATCCAATAAGGACACACCTAGGAGTAAAGGAGAAGAAC
Sen1-1907-GFP-r-AvrII	GTTCTTCTCCTTTACTCCTAGGTGTGTCCTTATTGGATTTTACCTC
Cloning	
SEN1 upstream Sal1	TTTTTGTCGACTTTAGTTAAAACGCGTTACATAG
SEN1 mid-down PstI	TGCTAGTTCTTCTGCAGAATCCAGGGCAG
SEN1 mid-up PstI	CTGCCCTGGATTCTGCAGAAGAACTAGCA
SEN1 downstream	TTTTTACTAGTAATGCAAATAGCTAGAGTAGCC

Spe1	
Sen1 5'-PstI(start)	CAACTGCAGAATTCATTTATAATAAACAGATGCGCTTT
SEN1-1B-upstream	CATGAAGCCTTTACTACTTCTAGAATCTTGG
SEN1-1B-downstream	CAGCCCTTCCAGTGTTGAATACTC
SEN1-1C-upstream	TTAAACCTCAATTGGTCCGTGTAGG
SEN1-1C-downstream	ATACCCATTGTTGCGAGTACATCATG
Sen1 1089-PstI fwd	CTTCTGCAGTCAAGAGACAAACTGCGGCT
Sen1 1103-PstI fwd	CTTCTGCAGAACATATGAGAAAGAGATTAAATGTTGAT
Sen1 1135-PstI fwd	CTTCTGCAGAGCCTATTGGTAATTATTCTGACG
Sen1 1151-PstI fwd	CTTCTGCAGACTACCAAAAAGTCATGAAGC
AvrII-GFP-S65T-fwd	TTTCCTAGGAGTAAAGGAGAAGAACTTTTCACTGGAG
AvrII-GFP-S65T-rev	TTTCCTAGGTTTGTATAGTTCATCCATGCCATGTG
PRS313-UP-BSPEI	AAAAATCCGGAATTCATTTATAATAAACAGATGCGCTTTTTTATATATATAG
PRS313-DOWN-BSPEI	AAAAATCCGGATCTTGAAGTAAAAGTGTTAGGAATTATACCTG
SETX 5'-BspEI	AAAAATCCGGAACATGTTGTTGGTGTACGCCAG
SETX 3'-BspEI	AAAAATCCGGACTATAAAAGCTTTCTTTTCTTGGAACTGCTGTC
SETX Ncol down	TTTTTCCATGGACCCCTCTGGGGCTATGGTAGG
SETX BspE1 up	TTTTTTCCGGAGTTACCTTCTCATGTTCAGGCG
SETX BspE1 down	TTTTTTCCGGATCCATTCTGTGGTTTACTTGGCTG
SETX Pml1 up	TTTTTCACGTGATAGAGAGAATTTCTATCAGGTGC
SETX Bsg1 up	TTTTTGTGCAGCACAGGAGTATGGCTACGACCA
YeHu 1B up(SphI)	CTTGCATGCGAGAGAATTTCTATCAGTTGCAAGTACG
YeHu 1B down(ShpI)	CTTGCATGCTCTTCCTTTGTGTAGTTACCAGAGAAC
YeHu 1C up(SphI)	CTTGCATGCCTAAAGTTCAGTTTGGACAGCCAAG
YeHu 1C down(SphI)	ATTGCATGCGATATGGGACTCTAAGATGATGATACTCTG

SETX 1C BsrGI 5'-end	AAAATGTACAGGACAGCCAAGTAAACCACAG
SETX 1C BsrGI 3'-end	AAAATGTACAAGGATGATGATACTCTGTGTTTTCTGTGG
pRS313 1B 5'-SphI	CAAGCATGCGTGAAGAACATAGACCTTGCCA
pRS313 1B 3'-SphI	CAAGCATGCAAATGACGACTATTGAAAGAGAGT
pRS313 1C 5'-SphI	CAAGCATGCAACGTTTACAACGTCTGACC
pRS313 1C 3'-SphI	CAAGCATGCTCGACTTTATCTGGTTCAGCCCA
Del sen1-1C BsrG1 fwd	GTAAACGTTGCAATTAAGGACCTTGTACAAGACATCATATGTTCGACTTTATCTG
Del sen1-1C BsrG1 rev	CAGATAAAGTCGAACAATTGATGTCTTGTACAAGGTCCTTAATTGCAACGTTTAC
GFP(S65T)-1S-fwd	AGTAAAGGAGAAGAACTTTTCACTGGAG
SEN1-1905K-FWD	AAGGACACAAAGAAGAGAAGAGTTGTC
SEN1-1073K-FWD	AAGGCAAAAGGCAAAGGAATACA
SEN1-1929E-REV	CTCTTTCTTCTTTTTCTTTACAGCTTTATCTGC
SEN1-N2-REV	ATTCATTTATAATAAACAGATGCGCTTTTTTATA
SEN1-1072A-REV	AGCAATAGCAAAAAGCTCCCGAG
SEN1-1930K-FWD	AAGAAGAAGGAAAAGAAAAAGTCCAAGG
SEN1-1091R-REV	TCTCTTGACAACCTTACCATTAATATCGAC
Nab3_Q777_R	CTGTTGTTGTTGTTGTTGTTGTTG
Nab3_stop_F	TAGACTCCCTTTTTTCAATCTTTTCC
YDR042C up	GCCTTAAAATTTCTACCCTTTCTACTTTTCAAGG
YDR042C down	GAGTATTGCTCTTGCCATAAAAGTGTGTTC
SNR47(100-215) fwd	GGTCTCGAGATATATTTTCGCGTCATTCTTG
SNR47(100-215) rev	GGTCTCGAGCAGAAATAAAGAAAATGAAAGC
SNR47 upstream	ATGATGAATTCCTATAACAACAACAACATG
SNR47 downstream	TATCAGTGTTAAAAAGCTATTGTCAAAGT
NRD1 fwd	CGAGTTACAGGAAAGGAACC

NRD1 rev	TCTATTCCTTCTTGAGTTATATC
NAB3(-571)F	GCTATCTGGAACTTTTGATAAGTTGTATTC
NAB3(+2659)R	AAATAGCATTTGATAAGGAAGAAATAAAAG
NAB3(-500)F_BamHI	TATGGATCCTAATCTTCGCTACTTCAAGTTTC
NAB3(+2609)R_SalI	TATGTCGACAGTTTTTCGTTCTGAAAGAGATG
Sequencing	
SEN1-8-F	CCAACAATCCTGATAATAATAATAGCAACAAC
SEN1-174-R	CTCTTGAATATTTGGATTTGTGCCTTG
SEN1-547-F	GAAACTGCAATGTACGAGTGTCTC
SEN1-991-F	ACATTCAAATTTCCGTTAGAGCCA
SEN1-1451-F	CTATTAAGGAAAATGAAAGGGCAATGC
SEN1-1914-F	GCTGCTATCAACGAGATTAATGGA
SEN1-2383-F	GACTATTCTGAACTGGAAAATTTCACC
SEN1-5634-F	CGTACCATCAGAGCAAGAAGATGA
SEN1-6100-F	GGTTTGAAGGAGACGAGAACAATTTC
SEN1 seq1F	TTCTGTGCCGGCGTCACCATCCCT
SEN1 seq1R	ACTCTTGGTCTCCCAGTTATTGAAG
SEN1 seq2F	GTTGATATGAACCCTCTATATGAAA
SEN1 seq2R	GGTAGTCAGCAGGAGAGTTGAAGAA
SEN1 seq3F	ATCAGAGATATATTGTGTGAAGGTT
SEN1 seq3R	ATTTACGGGAGGCGATGGTTTAGCC
SEN1 seq4F	AAAATAGCTCCCCATACTTGCTGGA
SEN1 seq4R	GGTCTTTTGTTCAAGATATCCATAC
SEN1 seq 5R	ATCTGCCTCTTCACCTTCATCGACA
SEN1 4406 F	GTGGATAAAAGGATAGGTGAAAGGAA
SEN 4436 R	GTCACAGCATTATTAAATTTACGTT
Nab3(-120)_F	TCTTGAGCATCTTTATTAGATTCTGCAC

Nab3(+505)_F	GAACTTCGCCGTGAAACATTGG
Nab3(+2484)_R	TGCAAACCATATCAATCTACTTTCC
Northern blot probe	
SNR39B-1R	CAGCCCCGAAAGATGGTGTAATTTAAGTTGACAAC
CWP2-1R	AGCACCATTTTCAGTTTGTTGAGAGATAGTTTCGGTG
CWP1-1R	TTACAACAAGTAAGCAGCTGCGACCGCTAGAGCA
FLO1-1R	GTCTCAGCAGCTACAGTATTGGTAGTCGTTTCAGCAG
ARG3-1R	CAAATGTCGCACCGTTTCTCTCAGCAACTTTCTTTG
CWP1up-1F	GCAGATACCCCTCTTGACGGCAAACATTGTGTG
CWP1up-1R	CACACAATGTTTGCCGTCAAGAGGGGTATCTGC
CWP2up-1F	GGCTTCTTGTTCATCATTTTCAATTCTCTTCTTGCCATC
CWP2up-1R	GATGGCAAGAAGAGAATTGAAAATGATGAACAAGAAGCC
NDJ1-rev	GGATACTCCAGAATTGACGACATTAACCAAGGTG

TABLE 2.2 Plasmids used in this dissertation

Plasmid	Description	Source
pRS313-SEN1(PstI)	*SEN1* ORF cloned into pRS313, dam$^-$	CHEN *et al.* 2014
pGEM-SETX	*SETX* ORF cloned into pGEM	SURAWEERA *et al.* 2007
pRS315-SEN1(AvrII)	An internal *Avr*II site created by silent mutation of codon 2230 in pRS315-SEN1(PstI)	CHEN *et al.* 2014
pRS313-SEN1/SETXhelicase	*SEN1* codons 1175-1890 replaced with *SETX* codons 1769-2484 by *Pml*I and *Nco*I	CHEN *et al.* 2014
pRS313-SEN1/SETXRec1A	*SEN1* codons 1175-1474 replaced with *SETX* codons 1769-2069 by *Pml*I and *BspE*I	CHEN *et al.* 2014
pRS313-SEN1/SETXRec2A1	*SEN1* codons 1478-1889 replaced with *SETX* codons 2073-2483 by *BspE*I and *Nco*I	CHEN *et al.* 2014
pRS313-SEN1/SETXRec2A2	*SEN1* codons 1635-1889 replaced with *SETX* codons 2226-2483 by *Bsg*I and *Nco*I	CHEN *et al.* 2014
pRS313-nab3-1-777	NAB3 codons 1-777 cloned into pRS313 by *BamH*I and *Sal*I	This study
pRS313-nab3-Q778X	*nab3-Q778X* cloned into pRS313 by *BamH*I and *Sal*I	This study
pRS316-HRP1-CUP1	610 bp of sequence upstream of the *HRP1* start site fused with the *CUP1* ORF and 3'-UTR, and cloned into pRS316 using *BamH*I/*Xho*I sites	KUEHNER and BROW 2008
pRS316-NRD1-CUP1	532 bp of sequence upstream of the *NRD1* start site fused with the *CUP1* ORF and 3'-UTR, and cloned into pRS316 using *BamH*I/*Xho*I sites	KUEHNER and BROW 2008
pRS316-IMD2-CUP1	377 bp of sequence upstream of the *IMD2* start site fused with the *CUP1* ORF and 3'-UTR, and cloned into pRS316 using *BamH*I/*Xho*I sites with T-71A mutation	KUEHNER and BROW 2008
pRS316-SEN1-CUP1	434 bp of sequence upstream of the *SEN1* start site fused with the *CUP1* ORF and 3'-UTR, and cloned into pRS316 using *BamH*I/*Xho*I sites	This study
pGAC24-SNR47(fwd)	*SNR47* terminator (positions +100 to +215 relative to *SNR47* start site) cloned into pGAC24 using *Xho*I sites	CHEN *et al.* 2014
pGAC24-SNR47(rev)	Reversed sequence of *SNR47* terminator (positions +100 to +215 relative to *SNR47* start site) cloned into pGAC24 using *Xho*I sites	This study
pGAC24-CYC1	The minimal 83-bp *CYC1* 3'-processing element cloned into pGAC24 using *Xho*I sites	STEINMETZ and BROW 2003
pRS315-SEN1-GFP	GFP (S65T) sequence cloned into the C-terminal end of *SEN1* in pRS315-SEN1(AvrII) using AvrII sites	CHEN *et al.* 2014
pRS315-SEN1-GFP (dam$^-$)	pRS315-SEN1-GFP construct, dam$^-$	CHEN *et al.* 2014
pRS315-sen1-965-2231-GFP	Derived from pRS315-SEN1-GFP with codons 4-964 deleted	This study
pRS315-sen1-1089-2231-GFP	Derived from pRS315-SEN1-GFP with codons 4-1088 deleted	This study
pRS315-sen1-1-1907-GFP	Derived from pRS315-SEN1-GFP with codons 1908-2229 deleted	This study
pRS315-sen1-1-1072-GFP	Derived from pRS315-SEN1-GFP with codons 1073-2229 deleted	This study
pRS315-sen1-965-1907-GFP	Derived from pRS315-SEN1-GFP with codons 965-1907 retained	This study
pRS315-sen1-965-1984-GFP	Derived from pRS315-SEN1-GFP with codons 965-1984 retained	This study

pRS315-sen1-1004-1072-GFP	Derived from pRS315-SEN1-GFP with codons 1004-1072 retained	This study
pRS315-sen1-1004-1091-GFP	Derived from pRS315-SEN1-GFP with codons 1004-1091 retained	This study
pRS315-sen1-1004-1907-GFP	Derived from pRS315-SEN1-GFP with codons 1004-1907 retained	CHEN *et al.* 2014
pRS315-sen1-1051-1907-GFP	Derived from pRS315-SEN1-GFP with codons 1051-1907 retained	This study
pRS315-sen1-1089-1907-GFP	Derived from pRS315-SEN1-GFP with codons 1089-1907 retained	This study
pRS315-sen1-1089-1929-GFP	Derived from pRS315-SEN1-GFP with codons 1089-1929 retained	CHEN *et al.* 2014
pRS315-sen1-1089-1984-GFP	Derived from pRS315-SEN1-GFP with codons 1089-1984 retained	This study
pRS315-sen1-1905-1929-GFP	Derived from pRS315-SEN1-GFP with codons 1905-1929 retained	CHEN *et al.* 2014
pRS315-sen1-1905-2231-GFP	Derived from pRS315-SEN1-GFP with codons 4-1904 deleted	This study
pRS315-sen1-1930-2231-GFP	Derived from pRS315-SEN1-GFP with codons 4-1929 deleted	This study
pRS315-sen1-NLS-1089-1907-GFP	Derived from pRS315-SEN1-GFP with codons 1089-1907 retained and SV40 NLS added before codon 1089	CHEN *et al.* 2014
pRS315-sen1-W1166S-GFP	Derived from pRS315-SEN1-GFP with the W1166X mutation	This study
pRS315-sen1-T1819A-GFP	Derived from pRS315-SEN1-GFP with the T1819A mutation	This study
pRS315-sen1-R1820Q-GFP	Derived from pRS315-SEN1-GFP with the R1820Q mutation	This study
pRS313-sen1-W1166S	Derived from pRS313-SEN1(PstI) with the W1166S mutation	CHEN *et al.* 2014
pRS313-sen1-W1166X	Derived from pRS313-SEN1(PstI) with the W1166X mutation	This study
pRS313-sen1-I1370R	Derived from pRS313-SEN1(PstI) with the I1370R mutation	CHEN *et al.* 2014
pRS313-sen1-I1371F	Derived from pRS313-SEN1(PstI) with the I1371F mutation	CHEN *et al.* 2014
pRS313-sen1-C1409Y	Derived from pRS313-SEN1(PstI) with the C1409Y mutation	CHEN *et al.* 2014
pRS313-sen1-N1413S	Derived from pRS313-SEN1(PstI) with the N1413S mutation	CHEN *et al.* 2014
pRS313-sen1-R1641E	Derived from pRS313-SEN1(PstI) with the R1641E mutation	CHEN *et al.* 2014
pRS313-sen1-E1579K/R1641E	Derived from pRS313-SEN1(PstI) with the E1579K/R1641E mutations	CHEN *et al.* 2014
pRS313-sen1-F1767L	Derived from pRS313-SEN1(PstI) with the F1767L mutation	CHEN *et al.* 2014
pRS313-sen1-K1788E	Derived from pRS313-SEN1(PstI) with the K1788E mutation	CHEN *et al.* 2014
pRS313-sen1-T1819A	Derived from pRS313-SEN1(PstI) with the T1819A mutation	This study
pRS313-sen1-L1517F/F1742L	Derived from pRS313-SEN1(PstI) with the L1517F/F1742L mutation	This study
pRS313-sen1-K1761R/T1819A	Derived from pRS313-SEN1(PstI) with the K1761R/T1819A mutation	This study

pRS425-SEN1(PstI) (dam$^-$)	pRS425-SEN1(PstI), dam$^-$	CHEN *et al.* 2014
pRS425-sen1-W1166X	Derived from pRS425-SEN1(PstI) with the W1166X mutation	This study
pRS423-SEN1(PstI)		CHEN *et al.* 2014
pRS313-sen1-Δ1B	Derived from pRS313-SEN1(PstI), codons 1174-1285 replaced with an *Sph*I site	CHEN *et al.* 2014
pRS313-sen1-Δ1C	Derived from pRS313-SEN1(PstI), codons 1451-1565 replaced with an *Sph*I site	CHEN *et al.* 2014
pRS313-YeHu-1B	*SETX* codons 1770-1889 cloned into pRS313-sen1-Δ1B using *Sph*I sites	CHEN *et al.* 2014
pRS313-YeHu-1C	*SETX* codons 2057-2150 cloned into pRS313-sen1-Δ1C using *Sph*I sites	CHEN *et al.* 2014
pRS313-YeHu-1B-1C	*SEN1* codons 1174-1285 and 1455-1562 replaced with *SETX* codons 1770-1889 and 2062-2145 using an *Sph*I and a *BsrG*I digestion site respectively	CHEN *et al.* 2014
pRS313-YeHu-SETX	*SETX* codons 3-2678(stop) cloned into pRS313-SEN1ORFdel(BspEI) using *BspE*I sites	CHEN *et al.* 2014
pRS423-YeHu-SETX	The *SEN1-SETX* chimeric region of pRS313-YeHu-SETX cloned into pRS423	This study
pRS425-YeHu-SETX	The *SEN1-SETX* chimeric region of pRS313-YeHu-SETX cloned into pRS425	This study
pRS313-sen1-965-2231	Derived from pRS313-SEN1(PstI) with codons 4-964 deleted	CHEN *et al.* 2014
pRS313-sen1-965-2231 (dam$^-$)	Derived from pRS313-SEN1(PstI) with codons 4-964 deleted, dam$^-$	CHEN *et al.* 2014
pRS313-sen1-1004-2231	Derived from pRS313-SEN1(PstI) with codons 4-1003 deleted	CHEN *et al.* 2014
pRS313-sen1-1051-2231	Derived from pRS313-SEN1(PstI) with codons 4-1050 deleted	CHEN *et al.* 2014
pRS313-sen1-1089-2231	Derived from pRS313-SEN1(PstI) with codons 4-1088 deleted	CHEN *et al.* 2014
pRS313-sen1-1089-2231 (dam$^-$)	Derived from pRS313-SEN1(PstI) with codons 4-1089 deleted, dam$^-$	CHEN *et al.* 2014
pRS313-sen1-1104-2231	Derived from pRS313-SEN1(PstI) with codons 4-1103 deleted	CHEN *et al.* 2014
pRS313-sen1-1135-2231	Derived from pRS313-SEN1(PstI) with codons 4-1134 deleted	CHEN *et al.* 2014
pRS313-sen1-1152-2231	Derived from pRS313-SEN1(PstI) with codons 4-1151 deleted	CHEN *et al.* 2014
pRS313-sen1-1-1858	Derived from pRS313-SEN1(PstI) with codons 1-1858 retained	CHEN *et al.* 2014
pRS313-sen1-1-1907	Derived from pRS313-SEN1(PstI) with codons 1-1907 retained	CHEN *et al.* 2014
pRS313-sen1-1-1984	Derived from pRS313-SEN1(PstI) with codons 1-1984 retained	CHEN *et al.* 2014
pRS313-sen1-1-2097	Derived from pRS313-SEN1(PstI) with codons 1-2097 retained	CHEN *et al.* 2014
pRS315-NLS-1089-1907	Derived from pRS315-sen1-NLS-1089-1907-GFP, self-ligated after excising the GFP sequence by AvrII digestion	CHEN *et al.* 2014
pRS313-sen1-965-1907	Derived from pRS313-SEN1(PstI) with codons 965-1907 retained	CHEN *et al.* 2014
pRS313-sen1-965-1984	Derived from pRS313-SEN1(PstI) with codons 965-1984 retained	CHEN *et al.* 2014

pRS313-sen1-1004-1907	Derived from pRS313-SEN1(PstI) with codons 1004-1907 retained	CHEN *et al.* 2014
pRS313-sen1-1050-1907	Derived from pRS313-SEN1(PstI) with codons 1050-1907 retained	CHEN *et al.* 2014
pRS313-sen1-1089-1907	Derived from pRS313-SEN1(PstI) with codons 1089-1907 retained	CHEN *et al.* 2014
pRS313-sen1-1089-1929	Derived from pRS313-SEN1(PstI) with codons 1089-1929 retained	CHEN *et al.* 2014
pRS313-sen1-1089-1984	Derived from pRS313-SEN1(PstI) with codons 1089-1984 retained	CHEN *et al.* 2014
pRS313-sen1-1089-2097	Derived from pRS313-SEN1(PstI) with codons 1089-2097 retained	CHEN *et al.* 2014
pRS313-sen1-1104-1907	Derived from pRS313-SEN1(PstI) with codons 1104-1907 retained	CHEN *et al.* 2014
pRS313*-SEN1(PstI)	The *Pst*I site in the backbone of pRS313-SEN1(PstI) mutated by Quikchange	This study

TABLE 2.3 Strains used in this dissertation

Strain	Genotype	Source
46**α**	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2*	LESSER and GUTHRIE 1993
46**a**	*MAT***a** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2*	LESSER and GUTHRIE 1993
DAB206	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS316-*SEN1*]	CHEN *et al.* 2014
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	Research Genetics
XCY361	*MAT***α** *his3-Δ1 leu2-Δ0 lys2-Δ0 ura3-Δ0 NHP6A-TagRFP-T:nat*	CHEN *et al.* 2014
nrd2-1	*MAT***a** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1-E1597K*	STEINMETZ and BROW 1996
nrd1-5	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nrd1-V368G*	STEINMETZ and BROW 1996
nab3-11	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS313-*nab3-11(F371L/P374T)*]	KUEHNER and BROW 2008
ssu72-G33A	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 ssu72-G33A*	STEINMETZ and BROW 2003
rpb11-E108G	*MAT***a** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 rpb11-E108G*	STEINMETZ *et al.* 2006a
hrp1-5	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 hrp1::HIS3* [pRS315-*hrp1-5(L205S)*]	KUEHNER and BROW 2008
XCY377	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nrd1-V368G nab3::KanMX* [pRS313-*nab3-11(F371L/P374T)*]	This study
YSC1021-669432	*MAT***α**/*MAT***a** *his3/his3 leu2/leu2 lys2/LYS2 MET15/met15 ura3/ura3 NAB3/nab3::KANMX*	Open Biosystems
XCY373	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nrd1-V368G nab3::KanMX* [pRS316-*NAB3*]	This study
JNK004	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS316-*NAB3*]	KUEHNER and BROW 2008
XCY389	*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS313-*nab3-Q778X*]	This study
XCY387	*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS313-*nab3-1-777*]	This study
XCY029	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*SEN1*]	CHEN *et al.* 2014
XCY298	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS425-*SEN1*]	This study
XCY250	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-965-2231*]	CHEN *et al.* 2014
XCY258	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-1004-2231*]	CHEN *et al.* 2014
XCY262	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-1089-2231*]	CHEN *et al.* 2014
XCY294	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-1-1907*]	CHEN *et al.* 2014

XCY311	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-1-1858*]	CHEN *et al.* 2014
XCY363	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-1004-1907*]	CHEN *et al.* 2014
XCY365	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-1089-1929*]	CHEN *et al.* 2014
XCY037	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-E1597K*]	CHEN *et al.* 2014
XCY033	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-R1641E*]	CHEN *et al.* 2014
XCY035	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-E1597K/R1641E*]	CHEN *et al.* 2014
UMY024	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-G1747D*]	CHEN *et al.* 2014
XCY141	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-T1819A*]	CHEN *et al.* 2014
XCY383	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-I1370R*]	CHEN *et al.* 2014
XCY278	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-W1166S*]	CHEN *et al.* 2014
XCY279	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-I1371F*]	CHEN *et al.* 2014
XCY167	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-N1413S*]	CHEN *et al.* 2014
UMY070	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-T1779P*]	CHEN *et al.* 2014
XCY031	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-K1788E*]	CHEN *et al.* 2014
XCY280	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-F1767L*]	CHEN *et al.* 2014
UMY021	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-R1850H*]	CHEN *et al.* 2014
XCY296	*MAT***α** *cup1Δ ura3 his3 trp1 lys2 ade2 leu2 sen1Δ3::TRP1* [pRS313-*sen1-W1166X*]	CHEN *et al.* 2014

CHAPTER 3: IDENTIFYING THE MINIMAL ESSENTIAL REGION OF SEN1 AND THE EFFECTS CAUSED BY AOA2 DISEASE MUTATIONS ON SEN1-DEPENDENT TERMINATION

This chapter is adapted from a research article published in *Genetics* (Chen *et al.*, 2014). Ulrika Müller created the DAB206 strain and helped make point mutations in *SEN1*, Kaitling Sundling created the NHP6A-TagRFP-T/*nat* construct and I performed all the other studies.

3.1 Introduction

To allow facile analysis of *SEN1* and its mutant alleles *in vivo*, our lab created the plasmid-shuffle yeast strain DAB206 (see Section 2.2). In this chapter, I describe using the DAB206 strain to determine that the minimum essential region of Sen1 is the helicase domain plus a flanking nuclear localization signal. Guided by the Upf1 structure model, I showed that the molecular defect underlying the *nrd2-1* mutation (E1597K) of Sen1 is disruption of an intradomain ionic bond with R1641. My studies strongly support the importance of the helicase domain in Sen1-dependent termination of Pol II transcription.

The tractable genome, short generation time and facile genetic engineering tools available make yeast ideal for modeling human diseases. In addition to the molecular mechanisms of Sen1-dependent termination, the yeast genetic system can also provide insights into the pathogenesis of the neurodegenerative diseases co-segregating with mutations in *SETX*, the human ortholog of yeast *SEN1*. As Senataxin has been implicated in Pol II transcription termination, one possible cause of the diseases is transcription termination defects due to dysfunction of Senataxin and consequent changes in gene expression (SKOURTI-STATHAKI *et al.* 2011). In an attempt to assess effects of the neurodegenerative disease mutations in yeast, I created the ones from the Senataxin helicase domain in Sen1with help from Ulrika Müller, analyzed their effects in yeast, and found they cause a broad range of effects, including termination defects.

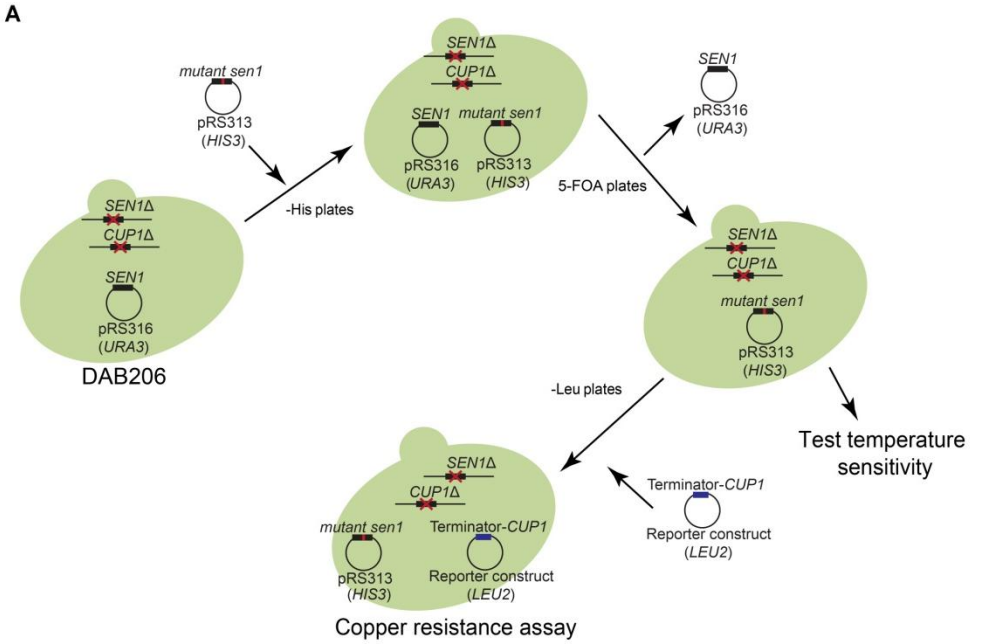
3.2 The essential region of Sen1 corresponds closely to the helicase domain

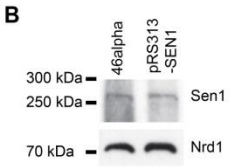
Previous genetic studies of *S. cerevisiae SEN1* used primarily chromosomal mutations and tested a small number of alleles (WINEY and CULBERTSON 1988; STEINMETZ and BROW 1996; FINKEL *et al.* 2010). The essential chromosomal *SEN1* gene was deleted previously

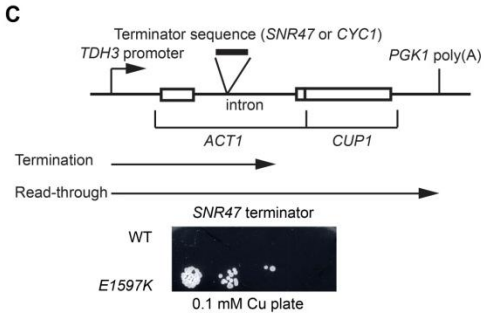
(DEMARINI *et al.* 1992), but the complementing plasmid used, although functional, was subsequently found to lack the *SEN1* promoter, 5'-UTR, and part of the protein-coding region. To overcome the previous limitations and allow more facile analysis of mutant *SEN1* alleles, our lab created a yeast strain DAB206 bearing the wild-type *SEN1* allele on a URA3-marked low-copy plasmid, pRS316, that allows plasmid shuffle with a mutant allele (Figure 3.1A). I confirmed that the expression level of plasmid-borne *SEN1* in strain, DAB206, appears similar to the level of expression of chromosomal *SEN1* in its parental strain, 46α (Figure 3.1B). We chose 46α as the parent of DAB206 because its chromosomal *CUP1* genes are deleted (Lesser and Guthrie 1993), which allows the use of *CUP1* reporter constructs to measure the extent of Sen1-dependent terminator read-through (Figures 3.1A, C and D). Although the *CUP1* reporter assay itself cannot distinguish increased terminator read-through from stabilization of a constitutive read-through transcript, we previously used nuclear run-on and/or Pol II chromatin immunoprecipitation to show that Sen1-E1597K induces read-through of all three terminators (*CYC1*, *SNR47* and *NRD1*) used in the *CUP1* reporter assay (STEINMETZ *et al.* 2006b).

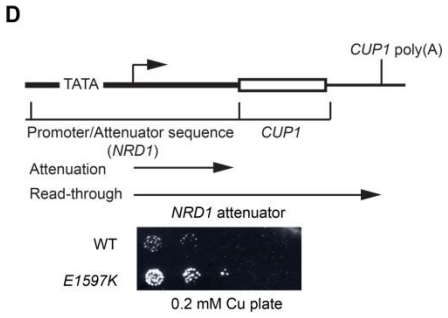
Since *SEN1* is essential in yeast, the first question I wanted to address is what part of Sen1 confers the required function. I used DAB206 to determine the minimal essential region of the Sen1 protein. Previously, a plasmid lacking the promoter, 5-UTR, and first 975 codons of the *SEN1* gene was found to complement a heat-sensitive mutation in *SEN1* (DEMARINI *et al.* 1992). I further delineated the essential region of Sen1 by systematically deleting the N- and C-terminal codons of the *SEN1* ORF while leaving the first three codons and stop codon intact. For most constructs, an additional alanine codon from the restriction digestion site is present between codon 3 and the first downstream codon. Mutant *SEN1* alleles were introduced into DAB206 in a *HIS3*-marked, low copy-number pRS313 vector. Recessive viability of each mutant was

FIGURE 3.1 A plasmid-based system for testing mutations in yeast *SEN1*. A) Schematic work pipeline for the plasmid-shuffle yeast strain. The chromosomal *SEN1* and *CUP1* genes of the strain DAB206 are disrupted, and the *SEN1* deletion is complemented by pRS316-SEN1. A plasmid-borne mutant *SEN1* allele is transformed into DAB206, and the spontaneous loss of pRS316-SEN1 is selected for by plating on medium containing 5-FOA. After transformation with a terminator-reporter construct, terminator read-through can be assayed by plating on medium containing copper sulfate, which is detoxified by the *CUP1* gene product. **B)** Anti-Sen1 helicase domain (HD) immunoblotting of cell extracts from strains with either chromosomal (46alpha) or plasmid-borne (pRS313-SEN1) *SEN1*. Nrd1 antiserum was used as a loading control (bottom). **C)** Schematic of the *ACT1-CUP1* construct for measuring terminator activity. The wild-type and a read-through mutant *sen1-E1597K* bearing the *SNR47* terminator-*CUP1* reporter construct were assayed on medium containing 0.1 mM copper. **D)** Schematic of the *CUP1* reporter for measuring attenuator activity. The wild-type and a read-through mutant *sen1-E1597K* bearing the *NRD1* attenuator-*CUP1* reporter construct were assayed on medium containing 0.2 mM copper.


A
mutant sen1
pRS313
(HIS3)
SEN1Δ
CUP1Δ
SEN1
pRS316
(URA3)
DAB206
-His plates
SEN1Δ
CUP1Δ
SEN1 mutant sen1
pRS316
(URA3)
pRS313
(HIS3)
5-FOA plates
SEN1
pRS316
(URA3)
SEN1Δ
CUP1Δ
mutant sen1
pRS313
(HIS3)
-Leu plates
Test temperature
sensitivity
SEN1Δ
CUP1Δ
mutant sen1
Terminator-CUP1
pRS313
(HIS3)
Reporter construct
(LEU2)
Terminator-CUP1
Reporter construct
(LEU2)
Copper resistance assay


B
46alpha
pRS313
-SEN1
300 kDa
250 kDa
Sen1
70 kDa
Nrd1


C
Terminator sequence (SNR47 or CYC1)
TDH3 promoter
PGK1 poly(A)
intron
ACT1
CUP1
Termination
Read-through
SNR47 terminator
WT
E1597K
0.1 mM Cu plate


D
CUP1 poly(A)
TATA
Promoter/Attenuator sequence
(NRD1)
CUP1
Attenuation
Read-through
NRD1 attenuator
WT
E1597K
0.2 mM Cu plate

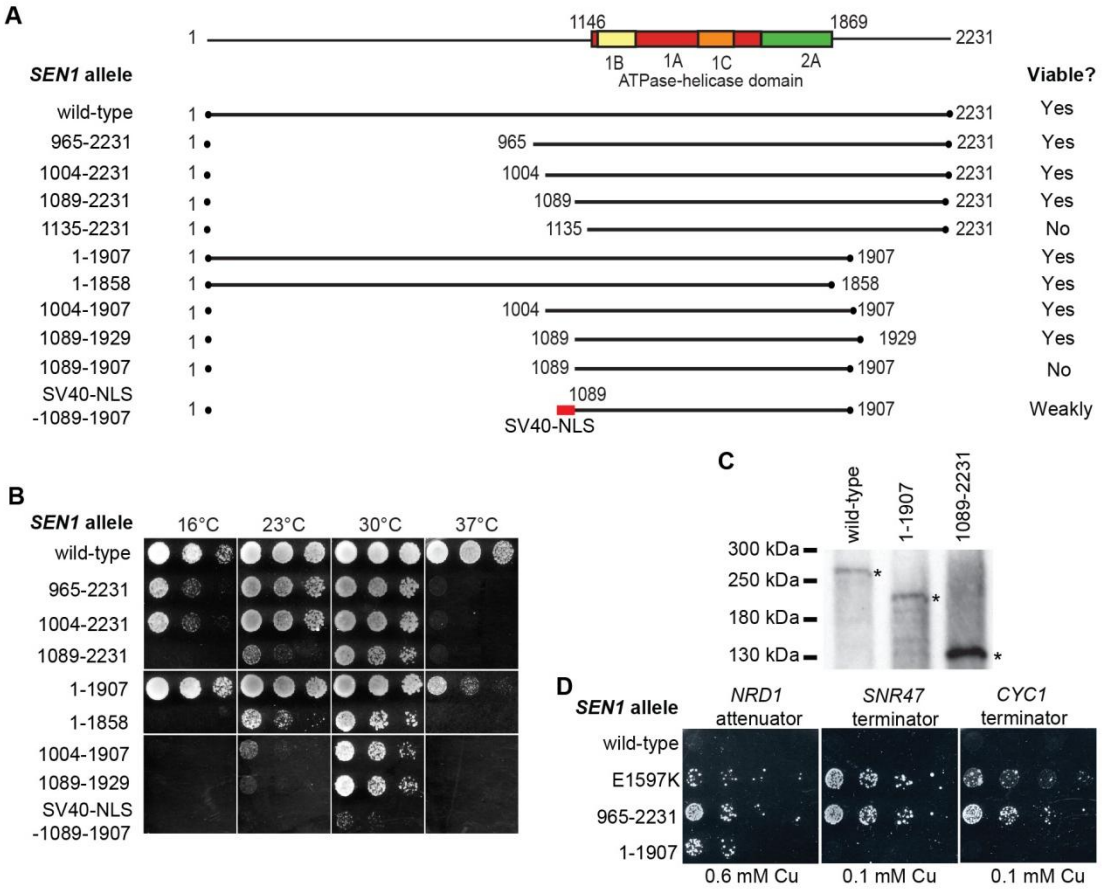
determined by plating transformants onto medium containing 5-FOA (Figure 3.2A). For viable *SEN1* mutants, their recessive growth phenotypes were scored at four temperatures: 16 ℃, 23 ℃, 30 ℃ and 37 ℃ (Figure 3.2B).

Deletions of codons 4-964 and 4-1003 resulted in strong heat-sensitivity and mild cold-sensitivity. N-terminal deletion through codon 1088 strongly increased the cold-sensitivity, whereas deletion through codon 1134 is recessive lethal at 30 ℃. Sen1 1089-2231 appeard to be over-expressed *in vivo* (Figure 3.2C), suggesting either that Sen1 levels are autoregulated or that the compromised function of the truncated protein at 30 ℃ selects for cells that have amplified the plasmid containing this allele. Deletion of the C-terminal 324 residues (codons 1908-2231) caused only weak heat-sensitivity at 37 ℃ (Figure 3.2B) and no obvious change in expression level (Figure 3.2C). However, deletion of an additional 49 C-terminal codons (to codon 1858), including the extreme C-terminus of the helicase domain, resulted in severe growth defects at all temperatures tested.

When N- and C-terminal codon deletions were combined, we identified the minimal length of the *SEN1* ORF for viability, which includes codons 1004-1907 or 1089-1929 (Figure 3.2A). Both of these alleles confer strong cold and heat sensitivity (Figure 3.2B). Thus, the essential region of Sen1 is ~900 amino acids long and corresponds closely to the helicase domain. However, a construct with only the 819 residues present in both viable constructs (codons 1089-1907; hereafter called "Sen1(HD)") is not sufficient, so an essential function is supplied by the either the N- or C-terminal residues immediately flanking the helicase domain.

I transformed the *CUP1* reporter constructs of the *CYC1*, *SNR47* and *NRD1* terminators into the viable deletion strains to measure the extent of terminator read-through. Analysis of copper sensitivity showed that viable deletions of the N- and C-terminal regions of Sen1 both

FIGURE 3.2 The essential region of *SEN1* closely corresponds to the helicase domain. A) Schematic of the *SEN1* alleles tested in this study and viability of the indicated alleles in the haploid strains. Numbers indicate the amino acid residues present. **B)** Recessive growth phenotypes of the truncated Sen1 alleles in a haploid strain. Eight-fold serial dilutions of cultures (starting from O.D. = 1.0) were spotted on YEPD plates and incubated at 16, 23, 30 and 37 ℃. **C)** Anti-Sen1(HD) immunoblot of cell extracts from haploid strains containing the indicated *SEN1* alleles. Bands with the expected apparent molecular mass are indicated with asterisks. **D)** Copper resistance assay for function of different terminators. Eight-fold serial dilutions of cultures (starting from O.D. = 1.0) of strains containing the indicated *SEN1* alleles and reporter genes were spotted on synthetic complete medium with copper sulfate at the indicated concentration and incubated at 30 ℃.


A
1146
1869
1
2231
1B
1A
1C
2A
ATPase-helicase domain
SEN1 allele
Viable?
wild-type
Yes
965-2231
Yes
1004-2231
Yes
1089-2231
Yes
1135-2231
No
1-1907
Yes
1-1858
Yes
1004-1907
Yes
1089-1929
Yes
1089-1907
No
SV40-NLS -1089-1907
Weakly
SV40-NLS
B
SEN1 allele
16°C
23°C
30°C
37°C
wild-type
965-2231
1004-2231
1089-2231
1-1907
1-1858
1004-1907
1089-1929
SV40-NLS -1089-1907
C
wild-type
1-1907
1089-2231
300 kDa
250 kDa
180 kDa
130 kDa
D
SEN1 allele
NRD1 attenuator
SNR47 terminator
CYC1 terminator
wild-type
E1597K
965-2231
1-1907
0.6 mM Cu
0.1 mM Cu
0.1 mM Cu

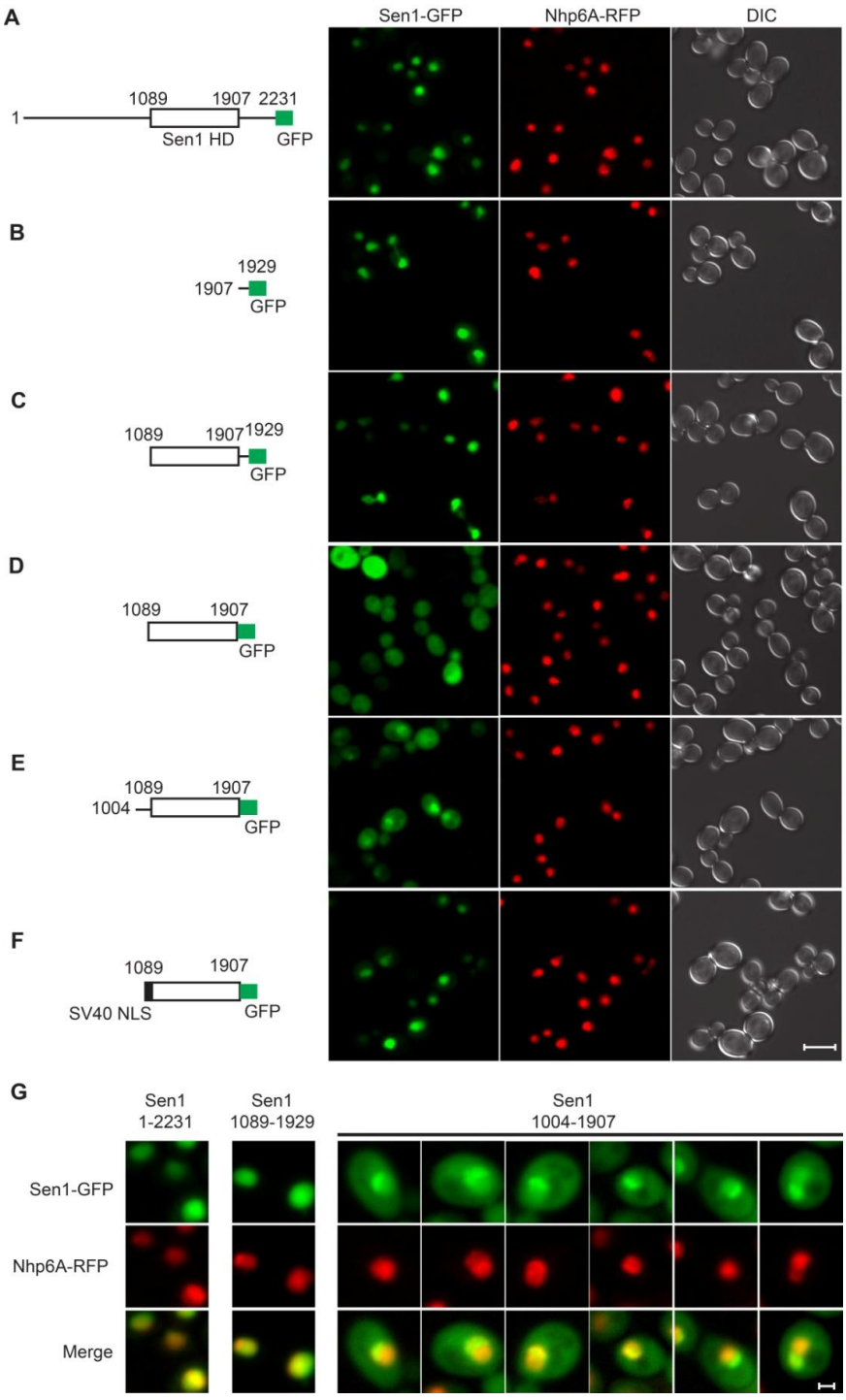
resulted in increased copper resistance indicating terminator read-through. The N-terminal deletion exhibited stronger read-through on all three terminators tested, consistent with its more severe high- and low-temperature growth defects (Figure 3.2D).

3.3 Nuclear localization of Sen1(HD) is required for cell viability

I sought to elucidate the essential function that is supplied by the residues flanking the helicase domain. Residues 1862-2093 of Sen1 were previously shown to contain a functional nuclear localization sequence (NLS) (URSIC *et al.* 1995), the N-terminal border of which was subsequently localized to residues 1890-1967 (NEDEA *et al.* 2008). Moreover, a bipartite NLS-like sequence spans Sen1 residues 1908-1929 (DEMARINI *et al.* 1992), which corresponds to the C-terminal essential flanking region (Figure 3.2A). I speculated that the flanking regions function redundantly in targeting the helicase domain to the nucleus. To test this hypothesis, I investigated localization of Sen1-GFP fusion proteins by fluorescent microscopy in the strain XCY361, which contains an intact chromosomal *SEN1* gene and RFP-labeled non-histone protein 6A (Nhp6A-RFP) as a nuclear marker (Figure 3.3).

Consistent with a previous study (URSIC *et al.* 1995), I found that wild-type Sen1-GFP predominantly localizes to the nucleus (Figure 3.3A). I further showed that the putative bipartite NLS is sufficient to direct nuclear localization, since GFP fused to Sen1 1907-1929 concentrated in the nucleus (Figure 3.3B). The subcellular distribution of Sen1(1089-1929)-GFP, which contains the bona fide NLS, resembles that of wild-type Sen1 (Figure 3.3C). Sen1(HD)-GFP, which contains only the helicase domain with no flanking sequences, displayed uniform distribution throughout the cell, with no nuclear enrichment (Figure 3.3D). However, Sen1(1004-1907)-GFP, which contains the helicase domain with N-terminal flanking sequence, exhibited weak nuclear localization (Figure 3.3E). This result indicates the N-terminal flanking sequence

FIGURE 3.3 N- and C-terminal flanking sequences localize the Sen1 helicase domain to the nucleus. A-F) Diagrams of the Sen1 constructs are shown at left. The micrographs show fluorescence of Sen1-GFP constructs (green) or Nhp6A-RFP (nuclear marker; red); differential interference contrast (DIC) images of the cells are shown at right. The scale bar for A-F is 10 μm. **G)** The images of representative cells show sub-nuclear localization of Sen1(1004-1907)-GFP (green) and Nhp6A-RFP (red). The incomplete co-localization of the two proteins is revealed in the merged images. Cells containing Sen1(1-2231)-GFP or Sen1(1089-1929)-GFP are shown on the left for comparison. The scale bar is 2 μm.


A
Sen1-GFP
Nhp6A-RFP
DIC
1
1089
1907
2231
Sen1 HD
GFP
B
1929
1907
GFP
C
1089
1907
1929
GFP
D
1089
1907
GFP
E
1089
1907
1004
GFP
F
1089
1907
SV40 NLS
GFP
G
Sen1
1-2231
Sen1
1089-1929
Sen1
1004-1907
Sen1-GFP
Nhp6A-RFP
Merge

has weaker NLS activity than the C-terminal NLS. Nevertheless, either the N- or C-terminal flanking sequence is able to direct sufficient Sen1 helicase core to the nucleus to support cell viability at 30℃.

To test if nuclear localization is the only essential function of the flanking sequences, I fused the SV40 NLS (PKKKRKVGIPQ) to the N-terminal end of Sen1(HD)-GFP. This fusion construct efficiently localized to the nucleus (Figure 3.3F). However, a strain containing the SV40 NLS-Sen1(HD) grew extremely slowly even at 30℃ (Figure 3.2B). Therefore, while the SV40 NLS can rescue Sen1(HD) function, it does so much less efficiently than either the N-terminal or C-terminal flanking sequences.

I noticed that the Sen1(1004-1907)-GFP signal and the Nhp6A-RFP signal tend to segregate from one another within the nucleus, with the former exhibiting predominantly a crescent-shaped distribution (Figure 3.3G). Such a distribution pattern was not observed with any of the other Sen1-GFP constructs, suggesting a subnuclear, perhaps nucleolar, localization signal exists in the N-terminal flanking region (see Chapter 3.8).

3.4 The *sen1-E1597K* substitution disrupts an intradomain salt bridge

Using a positive selection for Sen1-dependent terminator read-through, Steinmetz and Brow previously isolated the *nrd2-1* allele of *SEN1*, which harbors the missense mutation Glu1597 to Lys in the helicase domain (STEINMETZ and BROW 1996). This residue is adjacent to Motif II (Figure 3.1), which makes up part of the ATP-binding site, but the molecular defect of the E1597K mutation was unknown. Since no structural information is available for the Sen1 helicase domain, we used the yeast Upf1 helicase domain as a structure model (CHAKRABARTI *et al.* 2011). Based on sequence alignment (Figure 3.4), E1597 in Sen1 is equivalent to Upf1 E579, which forms a salt bridge with R623 in Upf1 that likely stabilizes the 1A domain (Figure 3.5A).

FIGURE 3.4 Sequence alignment between the S. cerevisiae Sen1 and Upf1 helicase domains. The identical and similar residues are highlighted in yellow. Substitutions of Sen1: L1517P, E1597K (*nrd2-1*), R1641E, F1742L, G1747D (*sen1-1*), K1761R and T1819A, are highlighted in red. E1597, R1641, G1747 and T1819 in Sen1 are equivalent to E579, R623, G727 and T800 in Upf1.

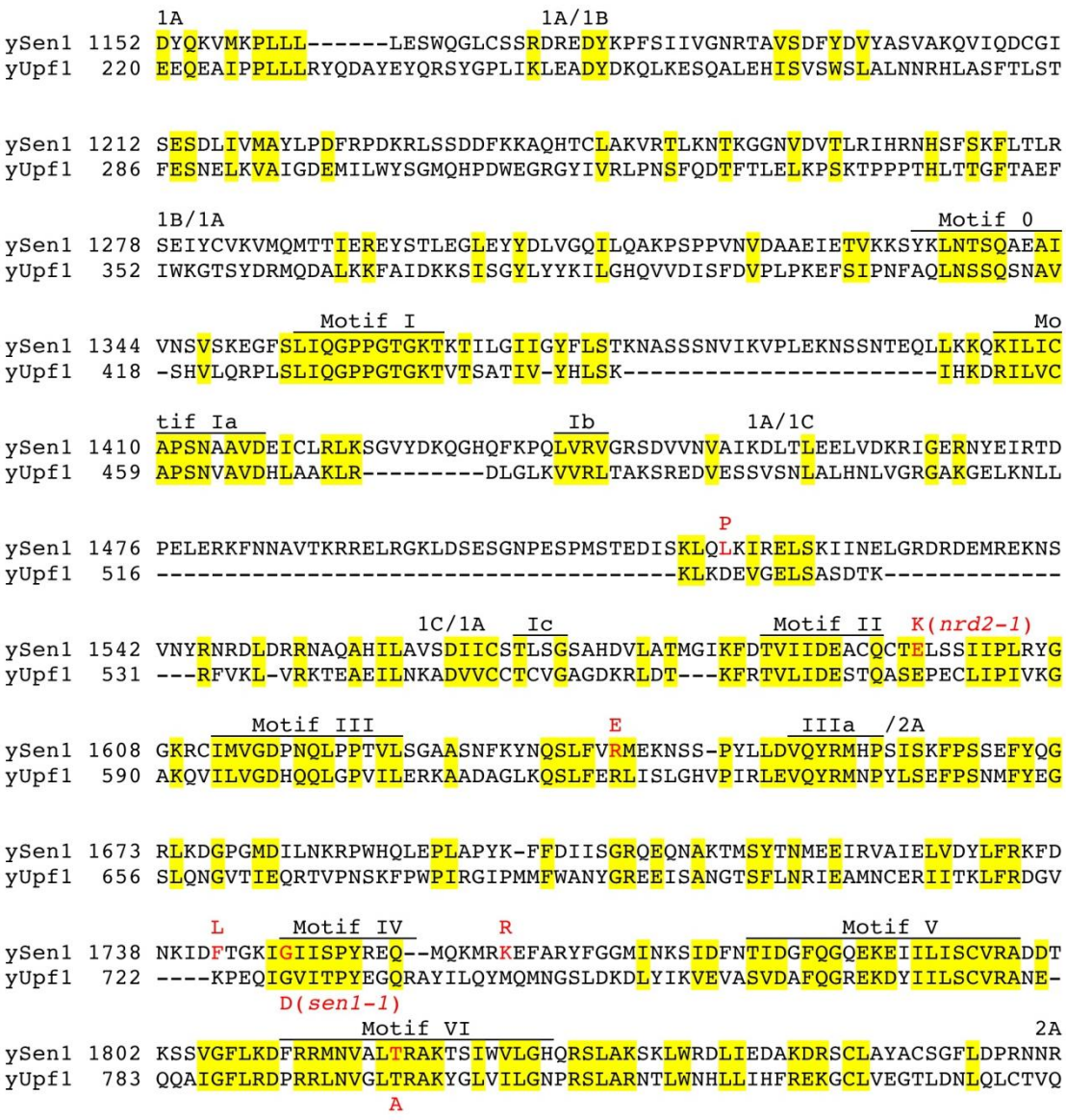
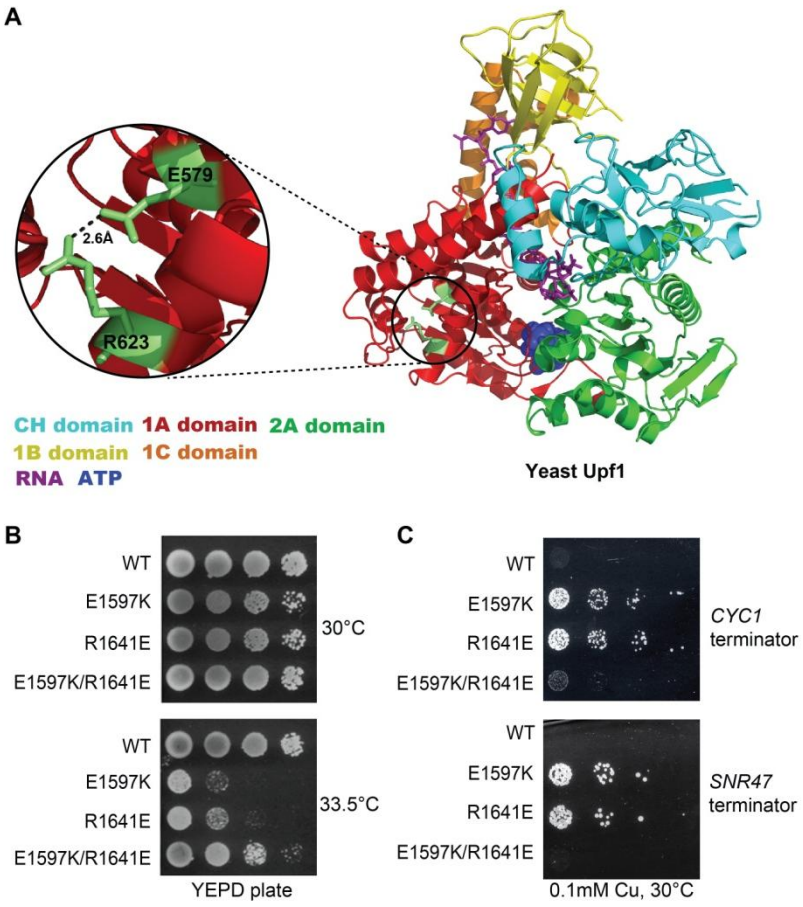

1A
1A/1B
ySen1 1152 DYQKVMKPLLL------LESWQGLCSSRDREDYKPFSIIVGNRTAVSDFYDVYASVAKQVIQDCGI
yUpf1 220 EEQEAIPPLLLRYQDAYEYQRSYGPLIKLEADYDKQLKESQALEHISVSWSLALNNRHLASFTLST
ySen1 1212 SESDLIVMAYLPDFRPDKRLSSDDFKKAQHTCLAKVRTLKNTKGGNVDVTLRIHRNHSFSKFLTLR
yUpf1 286 FESNELKVAIGDEMILWYSGMQHPDWEGRGYIVRLPNSFQDTFTLELKPSKTPPPTHLTTGFTAEF
1B/1A
Motif 0
ySen1 1278 SEIYCVKVMQMTTIEREYSTLEGLEYYDLVGQILQAKPSPPVNVDAAEIETVKKSYKLNTSQAEAI
yUpf1 352 IWKGTSYDRMQDALKKFAIDKKSISGYLYYKILGHQVVDISFDVPLPKEFSIPNFAQLNSSQSNAV
Motif I
Mo
ySen1 1344 VNSVSKEGFSLIQGPPGTGKTKTILGIIGYFLSTKNASSSNVIKVPLEKNSSNTEQLLKKQKILIC
yUpf1 418 -SHVLQRPLSLIQGPPGTGKTVTSATIV-YHLSK----------IHKDRILVC
tif Ia
Ib
1A/1C
ySen1 1410 APSNAAVDEICLRLKSGVYDKQGHQFKPQLVRVGRSDVVNVAIKDLTLEELVDKRIGERNYEIRTD
yUpf1 459 APSNVAVDHLAAKLR---------DLGLKVVRLTAKSREDVESSVSNLALHNLVGRGAKGELKNLL
P
ySen1 1476 PELERKFNNAVTKRRELRGKLDSESGNPESPMSTEDISKLQLKIRELSKIINELGRDRDEMREKNS
yUpf1 516 ----------KLKDEVGELSASDTK-------------
1C/1A
Ic
Motif II
K(nrd2-1)
ySen1 1542 VNYRNRDLDRRNAQAHILAVSDIICSTLSGSAHDVLATMGIKFDTVIIDEACQCTELSSIIPLRYG
yUpf1 531 ---RFVKL-VRKTEAEILNKADVVCCTCVGAGDKRLDT---KFRTVLIDESTQASEPECLIPIVKG
Motif III
E
IIIa
/2A
ySen1 1608 GKRCIMVGDPNQLPPTVLSGAASNFKYNQSLFVRMEKNSS-PYLLDVQYRMHPSISKFPSSEFYQG
yUpf1 590 AKQVILVGDHQQLGPVILERKAADAGLKQSLFERLISLGHVPIRLEVQYRMNPYLSEFPSNMFYEG
ySen1 1673 RLKDGPGMDILNKRPWHQLEPLAPYK-FFDIISGRQEQNAKTMSYTNMEEIRVAIELVDYLFRKFD
yUpf1 656 SLQNGVTIEQRTVPNSKFPWPIRGIPMMFWANYGREEISANGTSFLNRIEAMNCERIITKLFRDGV
L
Motif IV
R
Motif V
ySen1 1738 NKIDFTGKIGIISPYREQ--MQKMRKEFARYFGGMINKSIDFNTIDGFQGQEKEIILISCVRADDT
yUpf1 722 ----KPEQIGVITPYEGQRAYILQYMQMNGSLDKDLYIKVEVASVDAFQGREKDYIILSCVRANE-
D(sen1-1)
Motif VI
2A
ySen1 1802 KSSVGFLKDFRRMNVALTRAKTSIWVLGHQRSLAKSKLWRDLIEDAKDRSCLAYACSGFLDPRNNR
yUpf1 783 QQAIGFLRDPRRLNVGLTRAKYGLVILGNPRSLARNTLWNHLLIHFREKGCLVEGTLDNLQLCTVQ
A

FIGURE 3.5 The *sen1-E1597K* mutation disrupts an intradomain salt bridge. A) Predicted hydrogen bond between Upf1 residues E579 and R623 in the helicase domain crystal structure (PDB 2XZL, visualized by PyMOL, DeLano Scientific LLC). **B)** Growth phenotypes of strains containing the indicated *SEN1* alleles at permissive (30 ℃) and semi-permissive (33.5 ℃) temperatures after eight-fold serial dilution and spotting on YEPD plates. **C)** Copper resistance assay for terminator function. Strains containing the indicated *SEN1* alleles and either the *CYC1* or *SNR47* reporter construct were serially diluted eight-fold, spotted on synthetic complete medium with 0.1 mM copper sulfate and incubated at 30 ℃.


A
E579
2.6A
R623
CH domain 1A domain 2A domain
1B domain 1C domain
RNA ATP
Yeast Upf1
B
WT
E1597K
R1641E
E1597K/R1641E
30°C
33.5°C
YEPD plate
C
CYC1 terminator
SNR47 terminator
0.1mM Cu, 30°C

Upf1 R623 aligns with R1641 in Sen1, so this salt bridge may be conserved in Sen1.

To test whether the molecular defect of *sen1-E1597K* is disruption of the putative salt bridge, we introduced a R1641E substitution. If the proposed salt bridge exists, *sen1-R1641E* and *-E1597K* should exhibit the same defects by breaking this salt bridge, but in combination they should suppress such defects by mutually restoring the salt bridge. As shown in Figures 3.5B and 3.5C, R1641E alone conferred heat-sensitivity and terminator read-through phenotypes similar to those caused by E1597K. Moreover, the *E1597K/R1641E* double mutant was less heat-sensitive than either single mutant and exhibited no terminator read-through. These results support our hypothesis that the *nrd2-1* mutation disrupts a glutamate-arginine salt bridge between residues E1597 and R1641 that can be functionally replaced by a lysine-glutamate salt bridge. The success of this structure-guided mutational analysis demonstrates the utility of the Upf1 helicase domain structure as a predictive model for the Sen1 helicase domain structure, at least in the 1A and 2A domains, where the proteins are the most similar.

3.5 The human ortholog, *SETX*, cannot functionally replace yeast *SEN1*

Because mutations in *SETX* co-segregate with the neurodegenerative diseases AOA2 and ALS4, I wondered if the DAB206 strain can be used to study *SETX* and its relationship with the diseases. Initially, I attempted to replace yeast *SEN1* with human *SETX*, so that effects of the *SETX* disease mutations can be tested in yeast. I first tried to complement the *SEN1*-deletion in DAB206 with pRS313-SEN1/SETXORF or pRS313-SEN1/SETXhelicase, in which the entire protein-coding region or just the helicase domain of *SEN1*, respectively, is replaced with the corresponding regions of the *SETX* cDNA. However, the resulting strains did not grow on media containing 5-FOA, indicating that neither the entire *SETX* ORF nor its helicase domain can functionally replace their counterparts in *SEN1*, at least at the tested temperature of 30 ℃.

I further tested more SEN1/SETX chimeric constructs with the Sen1 helicase domain only partially replaced, since changing the entire helicase domain did not result in functional protein. The chimeric constructs include replacing *SEN1*'s Rec1A domain, Rec2A domain, insertion domain 1B, insertion domain 1C, and both insertion domains with their counterparts in *SETX*. However, none of these chimeric constructs render viability either. Therefore, even though the helicase domains are highly conserved between Sen1 and Senataxin, they may carry required sequences that have evolved differently between yeast and human.

3.6 AOA2 disease mutations in yeast *SEN1* cause growth defects and terminator read-through

Because replacing Sen1 with *SETX* or its chimeric constructs cannot render viability in yeast, I sought to characterize the disease mutations in the context of Sen1. This approach is restricted to the helicase domain as it is the only region where the sequence similarity is high enough to allow unambiguous identification of homologous residues. Among the 50 currently-identified missense mutations that co-segregate with AOA2 or ALS4, 28 are located in the Senataxin helicase domain (Table 3.1). Most AOA2 mutations in the helicase domain are in residues conserved in Sen1. In contrast, the two ALS4 mutations in the Senataxin helicase domain (R2136C and R2136H) are in a conserved residue that is not conserved in Sen1, so they cannot be tested in Sen1 (Figure 3.6).

I introduced 13 AOA2 mutations at analogous residues in Sen1 and characterized their effects *in vivo* (Figure 3.6, Table 3.2). Wild-type Sen1 and the two previously isolated mutations, E1597K and G1747D, were included for comparison. Plasmid-borne mutant *SEN1* alleles were transformed into the haploid yeast strain DAB206 to make merodiploid strains and observe their dominant effects. All 13 merodiploid strains were viable and grew normally at 30 ℃, indicating

TABLE 3.1 Previously reported missense mutations of ALS4 and AOA2

Location	Disease substitution	Disease
N-terminal domain	T3I (CHEN *et al.* 2004)	ALS4
	E65K (DUQUETTE *et al.* 2005)	AOA2
	M274V (DATTA and HOHLER 2013)	AOA2
	M274I (ASAKA *et al.* 2006)	AOA2
	W305C (MOREIRA *et al.* 2004; HAMMER *et al.* 2012)	AOA2
	P311L (BERNARD *et al.* 2008)	AOA2
	I331K (NANETTI *et al.* 2013)	AOA2
	R332W (MOREIRA *et al.* 2004)	AOA2
	L389S (CHEN *et al.* 2004; RUDNIK-SCHONEBORN *et al.* 2012)	ALS4
	P413L (MOREIRA *et al.* 2004)	AOA2
	H435R (FOGEL and PERLMAN 2006)	AOA2
	I466M (DAVIS *et al.* 2013)	AOA2
	H469R (CRISCUOLO *et al.* 2006)	AOA2
	P496L (NANETTI *et al.* 2013)	AOA2
	N563S (PILUSO *et al.* 2011)	AOA2
	A569C (GHROODA *et al.* 2012)	AOA2
	N603D (BASSUK *et al.* 2007)	AOA2
	Q653K (BASSUK *et al.* 2007)	AOA2
	V891A (RUDNIK-SCHONEBORN *et al.* 2012)	AOA2
	T918I (VANTAGGIATO *et al.* 2013)	AOA2
	R1294C (ASAKA *et al.* 2006)	AOA2
	P1503T (VANTAGGIATO *et al.* 2013)	AOA2
Helicase domain	F1756S (MOREIRA *et al.* 2004)	AOA2
	P1805S (ANHEIM *et al.* 2009)	AOA2
	A1945P (FOGEL and PERLMAN 2006)	AOA2
	L1976R (DUQUETTE *et al.* 2005; DATTA and HOHLER 2013)	AOA2
	L1977F (ASAKA *et al.* 2006)	AOA2
	C2006Y (ANHEIM *et al.* 2009)	AOA2
	N2010S (ANHEIM *et al.* 2009)	AOA2
	N2037D (FOGEL *ET AL.* 2014)	AOA2
	R2136H (CHEN *et al.* 2004)	ALS4
	R2136C (SAIGA *et al.* 2012)	ALS4
	L2155W (BERNARD *et al.* 2008)	AOA2

I2179S (FOGEL *et al.* 2014)	AOA2
H2197R (FOGEL and PERLMAN 2006; ANHEIM *et al.* 2009)	AOA2
D2207V (BERNARD *et al.* 2008)	AOA2
P2213L (MOREIRA *et al.* 2004; NANETTI *et al.* 2013)	AOA2
M2229T (NANETTI *et al.* 2013)	AOA2
I2264M (ANHEIM *et al.* 2009)	AOA2
F2363L (ANHEIM *et al.* 2009)	AOA2
P2368R (FOGEL and PERLMAN 2006)	AOA2
T2373P (BERNARD *et al.* 2008)	AOA2
R2380G (ANHEIM *et al.* 2009)	AOA2
R2380W (HAMMER *et al.* 2012)	AOA2
R2380Q (ANHEIM *et al.* 2009)	AOA2
K2382E (BERNARD *et al.* 2008)	AOA2
I2386T (ANHEIM *et al.* 2009)	AOA2
R2414Q (BERNARD *et al.* 2008)	AOA2
R2444H (BERNARD *et al.* 2008; ANHEIM *et al.* 2009; GAZULLA *et al.* 2010)	AOA2
R2444C (BERNARD *et al.* 2008)	AOA2

FIGURE 3.6 Sequence alignment between the Sen1 and Senataxin helicase domains. The identical and similar residues are highlighted in yellow. Senataxin disease mutant residues are indicated (AOA2: red, ALS4: green) with their disease substitutions below. The asterisks indicate the mutations tested in this study.

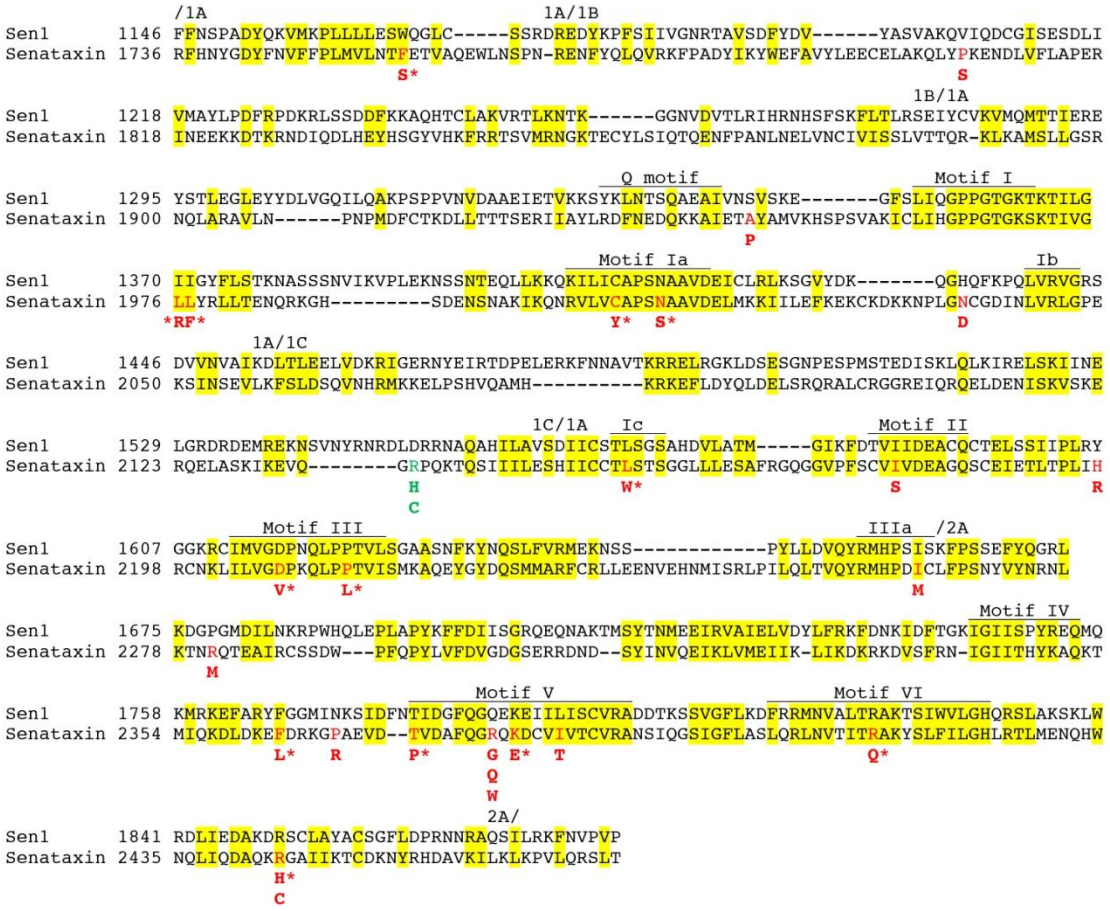

/1A
1A/1B
Sen1 1146 FFNSPADYQKVMKPLLLLESWQGLC-----SSRDREDYKPFSIIVGNRTAVSDFYDV------YASVAKQVIQDCGISESDLI
Senataxin 1736 RFHNYGDYFNVFFPLMVLNTFETVAQEWLNSPN-RENFYQLQVRKFPADYIKYWEFAVYLEECELAKQLYPKENDLVFLAPER
S*
S
1B/1A
Sen1 1218 VMAYLPDFRPDKRLSSDDFKKAQHTCLAKVRTLKNTK------GGNVDVTLRIHRNHSFSKFLTLRSEIYCVKVMQMTTIERE
Senataxin 1818 INEEKKDTKRNDIQDLHEYHSGYVHKFRRTSVMRNGKTECYLSIQTQENFPANLNELVNCIVISSLVTTQR-KLKAMSLLGSR
Q motif
Motif I
Sen1 1295 YSTLEGLEYYDLVGQILQAKPSPPVNVDAAEIETVKKSYKLNTSQAEAIVNSVSKE-------GFSLIQGPPGTGKTKTILG
Senataxin 1900 NQLARAVLN------PNPMDFCTKDLLTTTSERIIAYLRDFNEDQKKAIETAYAMVKHSPSVAKICLIHGPPGTGKSKTIVG
P
Motif Ia
Ib
Sen1 1370 IIGYFLSTKNASSSNVIKVPLEKNSSNTEQLLKKQKILICAPSNAAVDEICLRLKSGVYDK-------QGHQFKPQLVRVGRS
Senataxin 1976 LLYRLLTENQRKGH---------SDENSNAKIKQNRVLVCAPSNAAVDELMKKIILEFKEKCKDKKNPLGNCGDINLVRLGPE
RF
Y* S*
D
1A/1C
Sen1 1446 DVVNVAIKDLTLEELVDKRIGERNYEIRTDPELERKFNNAVTKRRELRGKLDSESGNPESPMSTEDISKLQLKIRELSKIINE
Senataxin 2050 KSINSEVLKFSLDSQVNHRMKKELPSHVQAMH----------KRKEFLDYQLDELSRQRALCRGGREIQRQELDENISKVSKE
1C/1A
Ic
Motif II
Sen1 1529 LGRDRDEMREKNSVNYRNRDLDRRNAQAHILAVSDIICSTLSGSAHDVLATM-----GIKFDTVIIDEACQCTELSSIIPLRY
Senataxin 2123 RQELASKIKEVQ--------GRPQKTQSIIILESHIICCTLSTSGGLLLESAFRGQGGVPFSCVIVDEAGQSCEIETLTPLIH
H
C
W*
S
R
Motif III
IIIa
/2A
Sen1 1607 GGKRCIMVGDPNQLPPTVLSGAASNFKYNQSLFVRMEKNSS----------PYLLDVQYRMHPSISKFPSSEFYQGRL
Senataxin 2198 RCNKLILVGDPKQLPPTVISMKAQEYGYDQSMMARFCRLLEENVEHNMISRLPILQLTVQYRMHPDICLFPSNYVYNRNL
V* L*
M
Motif IV
Sen1 1675 KDGPGMDILNKRPWHQLEPLAPYKFFDIISGRQEQNAKTMSYTNMEEIRVAIELVDYLFRKFDNKIDFTGKIGIISPYREQMQ
Senataxin 2278 KTNRQTEAIRCSSDW---PFQPYLVFDVGDGSERRDND--SYINVQEIKLVMEIIK-LIKDKRKDVSFRN-IGIITHYKAQKT
M
Motif V
Motif VI
Sen1 1758 KMRKEFARYFGGMINKSIDFNTIDGFQGQEKEIILISCVRADDTKSSVGFLKDFRRMNVALTRAKTSIWVLGHQRSLAKSKLW
Senataxin 2354 MIQKDLDKEFDRKGPAEVD--TVDAFQGRQKDCVIVTCVRANSIQGSIGFLASLQRLNVTITRAKYSLFILGHLRTLMENQHW
L* R P* G E* T
Q*
Q
W
2A/
Sen1 1841 RDLIEDAKDRSCLAYACSGFLDPRNNRAQSILRKFNVPVP
Senataxin 2435 NQLIQDAQKRGAIIKTCDKNYRHDAVKILKLKPVLQRSLT
H*
C

TABLE 3.2 Summary of the growth phenotype of the haploid AOA2 mutants and their read-through level

Mutation (Sen1)	Growth phenotype [a]	Mutation (Senataxin)	Max. viable Cu concentration (mM)[b]		
			NRD1	*SNR47*	*CYC1*
None	Normal	N/A	0.60	0.02	0.04
E1597K	H.S.	N/A	**1.20**	**0.60**	**0.20**
G1747D	H.S.	N/A	**1.00**	**0.30**	**0.30**
T1819A	H.S./C.S.	N/A	**1.20**	**1.00**	**0.80**
I1370R	Lethal [c]	L1976R	-	-	-
C1409Y	Lethal	C2006Y	-	-	-
L1569W	Lethal	L2155W	-	-	-
D1616V	Lethal	D2207V	-	-	-
P1622L	Lethal	P2213L	-	-	-
R1820Q	Lethal	R2414Q	-	-	-
W1166S	H.S.	F1756S	**1.00**	0.02	0.04
I1371F	H.S.	L1977F	0.50	0.02	0.02
N1413S	H.S.	N2010S	**0.80**	0.04	0.04
T1779P	H.S.	T2373P	**1.00**	**0.20**	0.04
K1788E	H.S.	K2382E	**1.20**	0.02	0.04
F1767L	Normal	F2363L	0.50	0.02	0.02
R1850H	Normal	R2444H	0.60	0.02	0.02

[a] H.S. = heat sensitive; C.S. = cold sensitive; Lethal = recessive lethal.

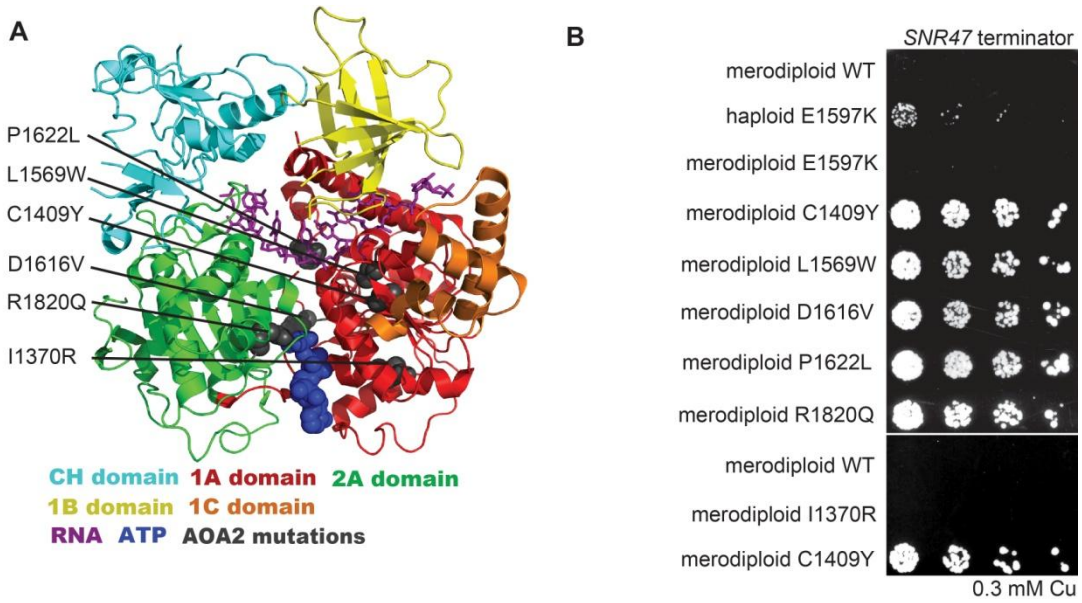
[b] The read-through level of each mutant is represented by its maximum viable copper concentration. Concentrations more than 0.1 mM higher than that of wild-type are considered as read-through and highlighted in underlined bold font.

[c] Viable haploid strain was obtained on medium containing 5-FOA at 23 ℃ but not 30 ℃.

that none of the disease alleles exhibit a dominant growth defect at this temperature. The merodiploid strains were then grown in medium containing uracil and plated on medium containing 5-FOA to select for clones that had lost the *URA3*-marked plasmid harboring the wild-type *SEN1* allele. Six of 13 AOA2 mutants (Table 3.2, Figure 3.7A) did not grow on medium containing 5-FOA at 30 ℃, indicating that these AOA2 mutations render Sen1 inactive for an essential function at this temperature. When these six alleles were tested for a dominant effect on read-through of the *SNR47* terminator, five exhibited snoRNA terminator read-through, with *sen1-I1370R* being the only exception (Figure 3.7B). Thus, these five non-functional AOA2 mutant Sen1 proteins inhibit the function of the wild-type protein on the *SNR47* terminator, and therefore must accumulate to an appreciable level. In contrast, given that the I1370R substitution replaces a buried hydrophobic residue with a charged residue, it could potentially result in protein misfolding and decreased thermostability, which would preclude a dominant effect on wild-type Sen1. Indeed, I found that the merodiploid strain with the I1370R substitution, but not the other recessively inviable ones, can grow on medium containing 5-FOA at 23 ℃, and the I1370R haploid mutant showed slow growth at 30 ℃ compared to other viable strains (Figure 3.7C). The growth defect of *I1370R* at 30 ℃ may be exacerbated on minimal medium containing 5-FOA, therefore prevented the haploid founder strain from growing under this condition.

All of the viable haploid mutants selected at 30 ℃ grew similarly to the wild-type Sen1 strain at 16 ℃, 23 ℃, and 30 ℃, but five of them displayed some degree of heat-sensitivity at 37 ℃ (Figures 3.8A and 3.8B). The W1166S substitution displays the strongest defect, followed by T1779P, then I1371F, N1413S and K1788E with minor growth defects. F1767L and R1850H exhibited no temperature sensitivity. In summary, the severity of the AOA2 mutations in Sen1 is variable. Of the 13 mutants we tested, six of them are recessive lethal and another five cause

FIGURE 3.7 A subset of AOA2-associated substitutions in Sen1 are recessive lethal and cause dominant terminator read-through. A) Locations of the 30 ℃ recessive lethal mutations modeled on the structure of the Upf1 helicase domain. **B)** Copper resistance assay for *SNR47* terminator function. Cultures of the merodiploid strains containing the chromosomal wild-type *SEN1* and the indicated plasmid-borne *SEN1* alleles were serially diluted eight-fold, spotted on synthetic complete medium with 0.3 mM copper sulfate, and incubated at 30 ℃. **C)** Growth phenotypes at 16, 23, 30 and 37 ℃ of the haploid wild-type and Sen1 mutants harboring the indicated substitutions after eight-fold serial dilution on YEPD medium.


A
P1622L
L1569W
C1409Y
D1616V
R1820Q
I1370R
CH domain 1A domain 2A domain
1B domain 1C domain
RNA ATP AOA2 mutations
B
SNR47 terminator
merodiploid WT
haploid E1597K
merodiploid E1597K
merodiploid C1409Y
merodiploid L1569W
merodiploid D1616V
merodiploid P1622L
merodiploid R1820Q
merodiploid WT
merodiploid I1370R
merodiploid C1409Y
0.3 mM Cu

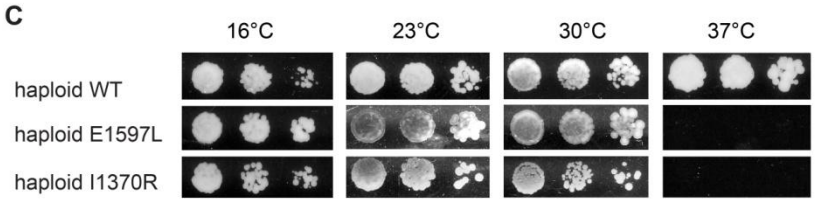
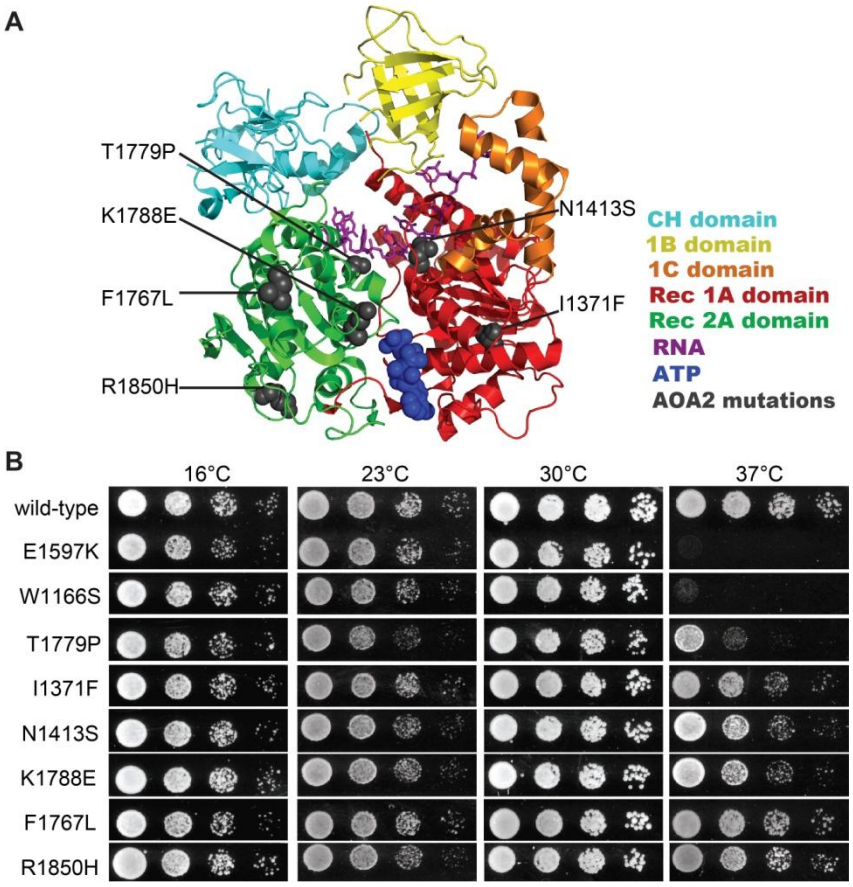

C
16°C
23°C
30°C
37°C
haploid WT
haploid E1597L
haploid I1370R

FIGURE 3.8 Some viable AOA2-associated substitutions in Sen1 confer heat-sensitivity. A) Locations of viable AOA2-associated substitutions modeled on the Upf1 helicase domain. The position of W1166 is not shown since a homologous residue could not be identified in Upf1. **B)** Growth phenotypes at 16, 23, 30 and 37℃ of the haploid Sen1 mutants harboring AOA2 substitutions after eight-fold serial dilution on YEPD medium.


A
T1779P
K1788E
F1767L
R1850H
N1413S
I1371F
CH domain
1B domain
1C domain
Rec 1A domain
Rec 2A domain
RNA
ATP
AOA2 mutations
B
16°C
23°C
30°C
37°C
wild-type
E1597K
W1166S
T1779P
I1371F
N1413S
K1788E
F1767L
R1850H

detectable heat-sensitivity.

We next tested if the viable AOA2 mutations cause defects in Sen1-dependent transcription termination. The haploid mutant strains were characterized in the copper resistance assay using the *SNR47, CYC1* and *NRD1* terminators. The viable mutants exhibited different levels of copper resistance (Table 3.2). Four out of five heat-sensitive AOA2 mutants, *sen1-W1166S, N1413S, T1779P* and *K1788E,* caused read-through of one or more terminators used in our assay, and the rest of the mutants showed no read-through effects. The terminators responded differently to the read-through mutants. The *NRD1* attenuator was responsive to all four AOA2 read-through mutants, with *sen1-N1413S* showing less read-through than the other three. Only *sen1-T1779P* caused *SNR47* terminator read-through, and no read-through of the *CYC1* terminator was observed in any of the AOA2 mutants. None of the AOA2 mutants caused read-through of *SNR47* or *CYC1* terminators to the levels of the previously isolated read-through mutant *sen1-E1597K*. In contrast, the founder *sen1-1* mutation, G1747D, exhibits read-through similar to the *nrd2-1* mutant, consistent with its defect in pre-tRNA splicing being an indirect effect of transcriptional read-through (WINEY and CULBERTSON 1988; STEINMETZ *et al.* 2001).

3.7 AOA2 mutations in Sen1 induce terminator read-through at the *NRD1* and *SNR47* loci

Since the *SNR47* terminator and *NRD1* attenuator are in an artificial context in their respective *CUP1* reporter genes, we sought to determine if termination defects also occur at the corresponding genomic loci in the presence of the AOA2 mutations. The strains carrying plasmid-borne wild-type *SEN1, sen1-E1597K, -W1166S, -T1779P* or *-K1788E* were chosen for analysis by Northern blotting. All the cultures were shifted from 30℃ to 37℃ for one hour before RNA extraction to enhance the defects from the mutants.

The Northern blot results are shown in Figure 3.9A and quantified in Figure 3.9B.

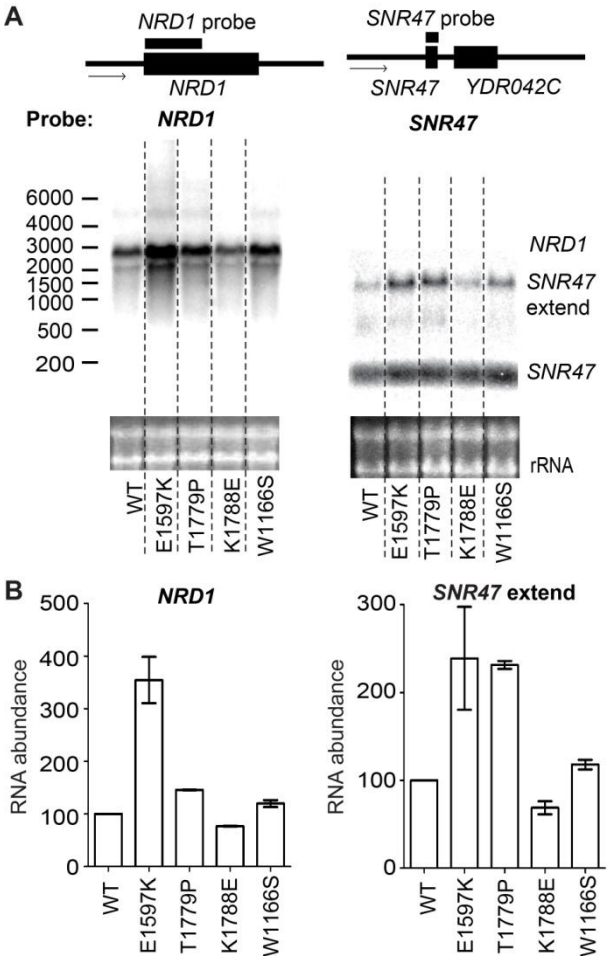
Previously, the chromosomal *nrd2-1* mutant (*sen1-E1597K*) was found to accumulate excess *NRD1* mRNA (due to read-through of the attenuator) and extended *SNR47* transcripts (STEINMETZ *et al.* 2001). We observed similar results with the plasmid-borne *sen1-E1597K* allele. Terminator read-through at these loci was also observed in some of the disease mutants. For the *NRD1* gene (Figures 3.9A and 3.9B, left panel), a small increase in transcripts in the *sen1-W1166S* and *-T1779P* strains implies decreased transcription attenuation. We did not observe excess *NRD1* mRNA in the *sen1-K1788E* strain despite its strong copper resistance in the presence of the NRD1/CUP1 reporter gene (Table 3.2). This inconsistency could be explained by the fact that the endogenous *NRD1* attenuator extends well into the Nrd1 coding region (ARIGO *et al.* 2006a), which is not included in the *NRD1/CUP1* fusion construct. K1788E may promote read-through of the truncated *NRD1* attenuator present in the reporter construct, but not the complete attenuator present at the endogenous *NRD1* locus.

For the *SNR47* locus (Figures 3.9A and 3.9B, right panel), accumulation of *SNR47-YDR042C* read-through transcripts were observed in the *sen1-W1166S* and *T1779P* mutants but not in *sen1-K1788E*. T1779P caused strong read-through of *SNR47* to a level similar to E1597K. Modestly elevated read-through of *SNR47* in the *sen1-W1166S* strain was detected in the Northern blot assay but not in the copper resistance assay, perhaps because this mutant protein is sensitive to increased temperature and RNA was prepared after a shift to 37℃, while the copper resistance assay was done at 30℃. In summary, our Northern blotting assays revealed termination defects caused by the AOA2 mutants at the endogenous loci, which in general agree with our *CUP1* reporter results.

3.8 Discussion

Several different functions in DNA and RNA transactions have been attributed to the

FIGURE 3.9 Terminator read-through of endogenous *NRD1* and *SNR47* transcripts in AOA2 mutant strains. A) Total RNA from haploid strains harboring the indicated *SEN1* alleles was resolved on denaturing agarose gels. The amplicons used as probes are indicated by the bars above the gene maps, and the arrows mark the transcription orientation. The numbers to the left of the blots represent the length in nucleotides of the RNA marker bands. The positions of the *NRD1*, *SNR47,* and *SNR47-YDR042C* read-through transcript (*SNR47*-extend) bands are indicated. Ethidium bromide stain of the portion of the filter containing the large and small ribosomal RNAs is shown below the blots. **B)** Quantification of the *NRD1* and *SNR47*-extended transcripts. The RNA abundance in each sample relative to the wild-type (100%) was calculated and averaged. Each experiment was performed twice with independent RNA extracts. The error bars indicate the range.


A
NRD1 probe
SNR47 probe
NRD1
SNR47
YDR042C
Probe:
NRD1
SNR47
6000
4000
3000
2000
1500
1000
500
200
NRD1
SNR47 extend
SNR47
rRNA
WT
E1597K
T1779P
K1788E
W1166S
B
NRD1
SNR47 extend
RNA abundance
500
400
300
200
100
0
300
200
100
0

essential yeast Sen1 protein and its human homolog, Senataxin (BROW 2011). In this chapter, I described *in vivo* studies of the structure-function relationships in Sen1 using our yeast genetic system, and address several open questions regarding the essential helicase domain of Sen1 and the effects caused by the human neurodegenerative disease mutations.

Multiple domains of Sen1 are involved in transcription termination

Only two conditional-lethal missense mutations (E1597K and G1747D) in conserved residues of the Sen1 helicase domain have previously been identified (DEMARINI *et al.* 1992; STEINMETZ and BROW 1996). My study has greatly expanded the pool of such missense mutations for elucidating the function of Sen1 and its helicase domain. Most of the missense mutations cause read-through of certain transcription terminators by Pol II, indicating the important role of the helicase domain in transcription termination.

My deletion analysis showed that nuclear localization of the Sen1 helicase domain is sufficient for viability, but that efficient transcription termination and normal growth at a range of temperatures also requires the N- and C-terminal domains of Sen1 (Figure 3.3D). The strong read-through effects of Sen1 965-2231 agree with the previous observation that the N-terminal part of Sen1 interacts with the Ser2-phosphorylated CTD, and impairment of this interaction by the mutation R302W decreases Sen1 occupancy on non-coding RNA genes (CHINCHILLA *et al.* 2012). Therefore, deleting the N-terminal region could dissociate Sen1 from the Pol II CTD, rendering it unable to efficiently recognize terminators. The C-terminal deletion of Sen1 (Sen1 1-1907) resulted in read-through of the *NRD1* attenuator. This could be caused by disrupting its interaction with Nab3 or other termination factors, such as the phosphatase Glc7 (NEDEA *et al.* 2008). However, I only observed moderate read-through in this deletion mutant strain, suggesting the interactions occurring in the C-terminal end are not critical for snoRNA

termination or are functionally redundant.

By tracking Sen1-GFP fusion proteins in merodiploid strains (Figure 3.4), we showed that either the N- or C-terminal flanking region can target the Sen1 helicase domain to the nucleus. The C-terminal flanking region contains a classical bipartite NLS that directs efficient import and is likely recognized by importin-alpha (Srp1) (MARFORI *et al.* 2011). The N-terminal flanking region does not contain a classical NLS and provides weak yet site-specific subnuclear localization activity. The resulting subnuclear distribution pattern resembles the shape of the nucleolus, which is not observed in the full length protein, suggesting a different mechanism for the localization activity of the N-terminal flanking region. The importance of this region in Sen1 has been revealed by our deletion analysis, as removing it strongly increased cold sensitivity (Figure 3.3B). We hypothesize that the N-terminal flanking region interacts with snoRNP proteins, which serves to bring it into the nucleus and localize it to the nucleolus. Such an interaction is presumably more transient in full-length Sen1, where it may serve to recruit Sen1 to co-transcriptionally assembling snoRNPs, increasing the efficiency of transcription termination on snoRNA genes (BALLARINO *et al.* 2005; YANG *et al.* 2005).

The heterologous SV40 monopartite NLS directs very efficient nuclear localization of the Sen1(HD), but results in very low viability. It is possible that a strong NLS at the N-terminus of the helicase domain interferes with helicase function in some way.

Terminator read-through is a common effect of the AOA2 mutations in Sen1

I tested thirteen AOA2 mutations from the Senataxin helicase domain in Sen1, and 77% (10 of 13) caused dominant or recessive terminator read-through (Figure 3.7B and Table 3.2), indicating that defects in transcription termination is a common feature of AOA2 mutations in yeast. The Upf1 helicase domain structure provides insights into the molecular processes

disrupted by the AOA2 mutations (Figures 3.7A and 3.8A). Some of the mutations that cause read-through localize to the catalytic core and are expected to disrupt binding of nucleic acid (N1413S and T1779P) or ATP (D1616V and R1820Q), or the coupling of these binding events (P1622L) (FAIRMAN-WILLIAMS *et al.* 2010). Other mutations are in the hydrophobic cores of the RecA domains (I1370R, I1371F, C1409Y, L1569W and K1788E), and may alter the overall fold of the helicase domain. No growth or termination defects were observed with the F1767L and R1850H AOA2 mutations. These residues are on the exterior surface of the helicase domain and may be involved in Senataxin-specific interactions.

In this study I examined the effects of *SEN1* mutations on Pol II termination. Functions of Sen1 in other nuclear processes have been proposed, including Pol I transcription termination, DNA repair, RNA processing and maintenance of genome stability (URSIC *et al.* 2004; KAWAUCHI *et al.* 2008; FINKEL *et al.* 2010; MISCHO *et al.* 2011; ALZU *et al.* 2012). Defects in the other processes may contribute to the growth defects we observed in the presence of the AOA2 substitutions, and our system can further be adapted to study Sen1 in those processes.

Potential of yeast for identifying Senataxin disease variants

I tried to develop the yeast system for studying human diseases. Given that 2 of the 13 AOA2 mutations that we tested in Sen1 cause no growth or termination defect, our yeast genetic system has a significant false negative rate in detecting disease-causing mutations in Senataxin. To determine the false positive rate, I would need to test single nucleotide polymorphisms (SNPs) in Senataxin that are known not to cause disease in humans and can be aligned accurately with the Sen1 sequence. However, no such SNPs have been identified.

I further wondered if there is a correlation between the yeast growth phenotypes and the severity of AOA2 mutations in human. As shown in Table 3.2, the tested AOA2 mutations were

categorized into "lethal", "heat sensitive" and "normal" groups based on their growth phenotypes in yeast. Using published data, I determined the average age of onset and the average serum alpha-fetoprotein (AFP) level for patients with the three observed combinations of alleles from these three phenotypic groups, namely two normal alleles, two lethal alleles, and one each of the heat-sensitive and lethal alleles (Table 3.3). For comparison, I also calculated average values for patients who have two presumptive null alleles (nonsense mutation, frameshift mutation, or a large deletion in the helicase domain). If the yeast phenotype were predictive of the human disease severity, I would expect the individuals with two "normal" alleles to have a significantly later age of onset and/or a lower serum AFP level. However, this analysis yielded no significant difference (using a t-test with $p < 0.05$) in either the age of onset or AFP level between any two of the yeast phenotype groupings. Therefore, although our yeast genetic system provides a means for identifying potential disease-causing sequence variants in Senataxin, the severity of AOA2 mutations in yeast shows no apparent correlation with severity in the AOA2 population for which we could access clinical data.

Notably, I found that five recessive lethal AOA2 mutations exhibit dominant read-through of the *SNR47* terminator (Figure 3.7B). Whether the corresponding *SETX* alleles cause any dominant effects in AOA2 carriers is not known. Although AOA2 is classified as a recessive disorder, our results in yeast suggest that a carrier trait may be detectable at the molecular level.

While the utility of our yeast genetic system for the characterization of the phenotypic consequences of human genetic variants is as yet uncertain, we have clearly established its utility for investigating structure-function relationships in yeast Sen1. More detailed analysis of the effects of the AOA2 mutations in Sen1, for example, by transcriptome studies, may yield insight into the nature of the molecular defects that lead to the symptoms of AOA2.

TABLE 3.3 Summary of the quantitative variables of the AOA2 patients

Patient group[a]	Average onset age ± S.D.[b]	Average AFP ± S.D.[b] (ng/mL)
Normal/Normal	17.7 ± 0.58 (3)	67.7 ± 83.40 (3)
H.S./Lethal	13.7 ± 4.94 (3)	78.2 ± 58.85 (3)
Lethal/Lethal	14.3 ± 3.87 (17)	19.3 ± 6.35 (3)
Null/Null	16.6 ± 4.63 (53)	41.3 ± 28.46 (37)

[a] Patients were grouped based on the yeast growth phenotype of each of their two *SETX* mutations; H.S. = heat sensitive; Lethal = recessive lethal.

[b] S.D. = Standard deviation; AFP = serum alpha-fetoprotein level.

CHAPTER 4: TRANSCRIPTOME-WIDE ANALYSIS OF THE TARGETS AND FUNCTIONAL RELATIONSHIPS OF SIX PROTEINS INVOLVED IN SEN1-DEPENDENT TERMINATION

Contribution:

The total RNA preparations from the mutant strains above and their parental wild-type strains, as well as the microarray hybridization and intial data analysis were performed by Kunal Poorey, Melissa Wells, Stefan Bekiranov and David Auble, University of Virginia Health System, Charlottesville (UVA). Raw microarray data were generated by the Microarray Core Facility at UVA. The same RNA preparations were used for the Northern blot analyses except Section 4.7. I performed all other analysis herein, with assistance from Ulrika Müller on some of the Northern blots.

4.1 Introduction

Efficient transcription termination by the Sen1 pathway relies on proper function of its termination factors. Thus, to understand how the Sen1 machinery terminates transcription, it is critical to identify the role of every component involved in this termination pathway and their functional relationships. Sen1-dependent termination factors have been shown to be differentially required for termination of different genes. In a small scale test of *HRP1*, *NRD1* and *IMD2* attenuators using a reporter assay, our lab found that variable subsets of Sen1-dependent termination factors are required for efficient termination of these attenuators (KUEHNER and BROW 2008). However, current studies of Sen1-dependent termination tend to ignore the differences between the termination factors. Without identifying the difference, the analyses of only one or two factors can hardly reach a reliable, comprehensive understanding of this pathway. To systematically understand the differential requirements for the Sen1 termination machinery on different genes, we expanded the study to the genome scale. By utilizing hypomorphic substitutions in the components of the Sen1 termination machinery, we sought to identify the effects of each component on the yeast transcriptome and their relationships.

As introduced in Chapter 1, substitutions in Sen1, Nrd1 and Nab3, which form the core Sen1 termination complex, and the Pol II C-terminal domain (CTD) Ser5 phosphatase Ssu72, were shown to cause Sen1-dependent termination read-through. Read-through mutations were also identified in three other proteins- Rpb3, Rpb11 and Hrp1. Two of them, Rpb3 and Rpb11, are the subunits of Pol II, and form a heterodimer located opposite the site of Pol II's DNA entry channel. The identified read-through substitutions in Rpb3 and Rpb11 cluster on a tiny portion on the surface of Pol II facing the nascent transcript (STEINMETZ *et al.* 2006a), suggesting the Rpb3-Rpb11 heterodimer may interact with Sen1 termination factors on the nascent transcript.

The nuclear RNA-binding protein Hrp1 is a subunit of cleavage factor I and thought to function primarily in cleavage and polyadenylation of the pre-mRNA 3' end, and mRNA export (KESSLER *et al.* 1997). A substitution in one of the two tandem RNA-recognition motifs (RRMs) of Hrp1 affects Sen1-dependent attenuation of the *HRP1* and *NRD1* genes, implicating Hrp1 as a Sen1-dependent termination factor, at least on *HRP1* and *NRD1* (KUEHNER and BROW 2008).

The six haploid mutant strains used in this chapter are *sen1-E1597K* (originally called *nrd2-1*), *nrd1-V368G* (*nrd1-5*), *nab3-F371L/P374T (nab3-11), ssu72-G33A, rpb11-E108G* and *hrp1-L205S (hrp1-5)*. The Nrd1, Nab3, Ssu72 and Hrp1 mutants are derived from the wild-type parental strain 46α, while the Sen1 and Rpb11 mutants are from 46a. The genotypes of the strains are listed in Section 2.2.

sen1-E1597K, nrd1-V368G, ssu72-G33A and *rpb11-E108G* were obtained from previous genetic selections by our lab, and they all exhibited defects in Sen1-dependent termination (STEINMETZ and BROW 1996; STEINMETZ and BROW 2003; STEINMETZ *et al.* 2006a). *nab3-F371L/P374T (nab3-11)* and *hrp1-L205S (hrp1-5)* were identified by Conrad *et al.* and Kessler *et al.*, respectively (KESSLER *et al.* 1997; CONRAD *et al.* 2000). Plasmids carrying the *NAB3* and *HRP1* mutant alleles were transformed into yeast, and both mutations were shown to cause read-through of certain Sen1-dependent terminators *in vivo* (KUEHNER and BROW 2008).

The E1597K substitution in Sen1 is located in its conserved helicase domain. It disrupts an intradomain ionic bond between E1597 and R1641 (see Section 3.4), and therefore the helicase-related activities. The RNA-binding proteins, Nrd1, Nab3 and Hrp1, are all essential for yeast cell viability, and substitutions in them are expected to interfere with their RNA-binding abilities. Nrd1 and Nab3 each has an RRM, and the read-through substitutions V368G and F371L/P374T are located in the RRMs of Nrd1 and Nab3 (Figures 4.1A and B), respectively.

Although Nrd1 and Nab3 may bind to RNA via ways other than the canonical RRM-RNA interaction, no RRM-independent RNA binding mechanism has been found in Nrd1 and Nab3 yet. In Nrd1, Bacikova *et al.* showed a helix-loop bundle domain of it and its RRM together can sandwich the RNA, which may provide an alternative way for Nrd1 and its RNA substrate to interact, but all the substitutions in the RRM they tested decreased the RNA binding affinity of the alternative binding mode (BACIKOVA *et al.* 2014). Hrp1 has two tandem RRMs, and the substitution L205S is located in one of them (Figure 4.1 C).

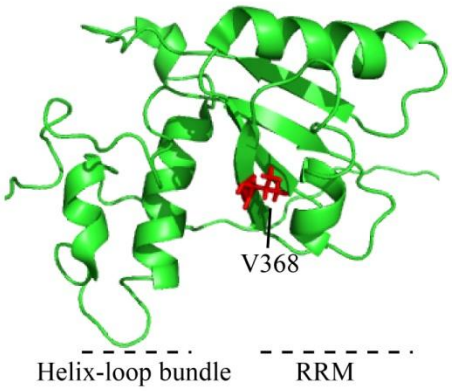
All the six mutant strains exhibit a heat-sensitive growth phenotype, and *rpb11-E108G* also shows cold-sensitivity at 16 ℃. In order to enhance the defects caused by these hypomorphic mutations for later RNA analyses, the Auble lab conducted a temperature shift on the 30 ℃ cultures of these mutants and their parent strains to 35 ℃ for 45 min before RNA extraction, except for the *rpb11-E108G* strain. This strain exhibits mild growth defects at 30 ℃, and its culture was not subjected to a temperature shift. A control culture of wild-type 46**a** strain was prepared in the same conditions as *rpb11-E108G*. The RNA extraction procedure is described in Section 2.5.

The raw microarray data were generated by the Microarray Core Facility at UVA. I processed the raw data to acquire the "absolute" and differential transcript level values of every sample. The details of data acquisition are described in Section 2.8. Since the microarray data are not strand-specific, I cannot discriminate transcripts and their changes of one strand from the other, which is a major limitation of this study.

By investigating the functional transcriptome data, I identified surprisingly diverse effects of the six components involved in Sen1-dependent termination. The *sen1-E1597K* substitution induces the strongest and broadest termination defects among the six proteins.

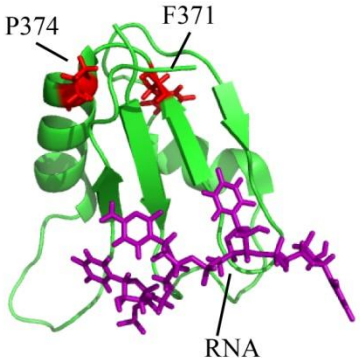
FIGURE 4.1 Location of terminator read-through substitutions in the structures of the RNA binding domains in *S. cerevisiae* Nrd1, Nab3 and Hrp1. The proteins are in green. The residues where the substitutions reside are highlighted in red. RNA is in purple. **A**) Structure of the Nrd1 helix-loop bundle and its RRM (BACIKOVA *et al.* 2014). **B**) Structure of the Nab3 RRM with RNA (HOBOR *et al.* 2011). **C**) Structure of the two tandem RRMs in Hrp1 with RNA (PEREZ-CANADILLAS 2006).

A



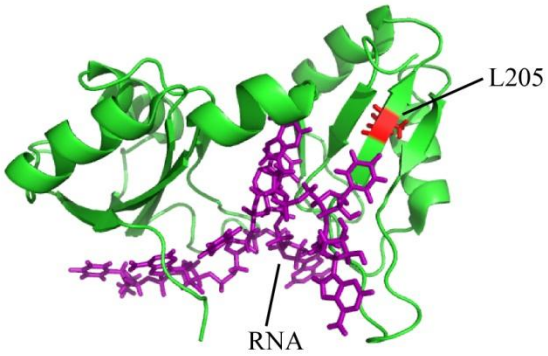
yeast Nrd1 helix-loop bundle and RRM (PDB: 2M88)

B



yeast Nab3 RRM (PDB: 2L41)

C



yeast Hrp1 2 RRMs (PDB: 2CJK)

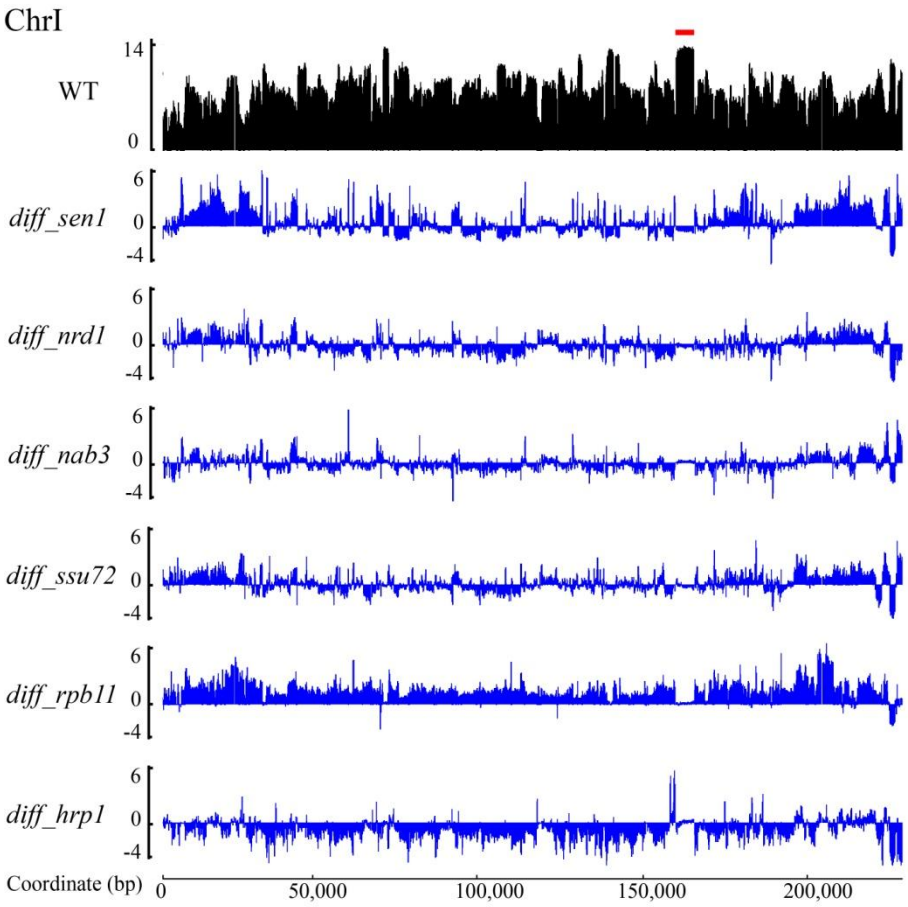
Pairwise correlation analyses show that transcriptome changes caused by substitutions in Sen1, Nrd1, Nab3 and Ssu72 are most highly correlated, consistent with the model that Sen1, Nrd1 and Nab3 form the core termination complex, and the recruitment of the complex is mediated by the phosphoralyted Pol II CTD. Sen1 appears to have a separate function that is different from Nrd1 and Nab3 but is affected by the substitution in Rpb11. Hrp1 exhibits a moderate degree of functional overlap with the Sen1 termination complex, but appears to be important for transcription processivity on protein-coding genes. I confirmed that termination of small nuclear (sn) and small nucleolar (sno) RNA genes are widely Sen1-dependent, but only a subset of these genes is affected by substitutions in Nrd1 and Nab3. I also showed the Sen1 substitution derepresses meiotic genes in vegetative cells, suggesting that Sen1 has an important role in regulating the meiotic transcriptome. Taken together, the results suggest that Sen1-dependent transcription termination is under complex, combinatorial control.

4.2 A substitution in the Sen1 helicase domain shows the highest magnitude read-through effects over the transcriptome

By visually inspecting the transcriptome profiles, I noticed that different mutants exhibited distinct changes in transcript level across the genome. As an example of transcript changes across the genome in different mutants, the overview of Chromosome I is shown in Figure 4.2. In brief, the Sen1 mutation induces transcript accumulation along major parts of this chromosome. The Nrd1, Nab3 and Ssu72 mutants exhibit almost equal amounts of transcript level increase and decrease. In contrast, the Rpb11 and Hrp1 mutations predominantly increase and decrease transcript levels of the entire chromosome, respectively.

In order to quantitatively evaluate the overall effects of each mutation on the transcriptome, I calculated the average differential transcript level across the genome of each

FIGURE 4.2 Overview of the transcript level of Chromosome I in wild-type and the differential transcript levels in the six mutants. The black trace shows the absolute transcript level in wild-type. The blue traces show the differential transcript levels in the indicated mutants. The y-axes represent the transcript levels in the log2 scale. The genome coordinate is indicated by the horizontal line at the bottom. The position of a transposable element is highlighted by a red bar.


ChrI
WT
14
0
diff_sen1
diff_nrd1
diff_nab3
diff_ssu72
diff_rpb11
diff_hrp1
6
0
-4
Coordinate (bp)
0
50,000
100,000
150,000
200,000

mutant in relative to the matched wild-type strain (Table 4.1). If a protein normally serves to restrict transcript accumulation (for example, by terminating transcription), I would expect to see an overall increase in transcript accumulation in a decreased function mutant. The magnitude of the change would reflect the importance of the protein in this process, the degree of the functional decrease due to the mutation, and the number of targets of the pathway across the genome.

The six mutants cause distinct changes on gene expression across the genome. The *sen1-E1597K* mutation causes the largest net increase in transcript accumulation genome-wide (23.4%), followed by *rpb11-E108G* (12.9%) and *ssu72-G33A* (5.4%). The *nab3-F371L/P374T* and *nrd1-V368G* mutations both resulted in a decrease of 8.7% and 11.8% respectively. In contrast, the *hrp1-L205S* mutation strongly decreased global transcript accumulation by 44.8%. I note that the magnitude of these changes may be underestimates due to the array normalization process. Nevertheless, according to the global analysis, Sen1 (or at least its function that is diminished by the mutation) generally restricts transcript accumulation, while Hrp1 functions to promote transcript accumulation. The other 4 proteins have less pronounced, but significant, effects on net transcript accumulation.

4.3 Substitutions in Sen1, Ssu72 and Hrp1 cause uneven changes of transcript level along the length of protein-coding genes

A gradual decrease of RNA level along the length of several protein-coding genes was observed in the Hrp1mutant by visual inspection. To see how the mutations affect transcript level of protein-coding genes in general, I performed metagene analysis to systematically characterize the transcript level changes on all protein-coding genes in the mutants. The coding region of every protein-coding gene was evenly divided into 10 segments, and I calculated the average

TABLE 4.1 Summary of the average gene expression fold changes across the genome of the mutants

Mutant	Mean of change (%)	95% confidence (%)
sen1	23.4	±0.1
rpb11	12.9	±0.1
ssu72	5.4	±0.1
nab3	-8.7	±0.1
nrd1	-11.8	±0.1
hrp1	-44.8	±0.1

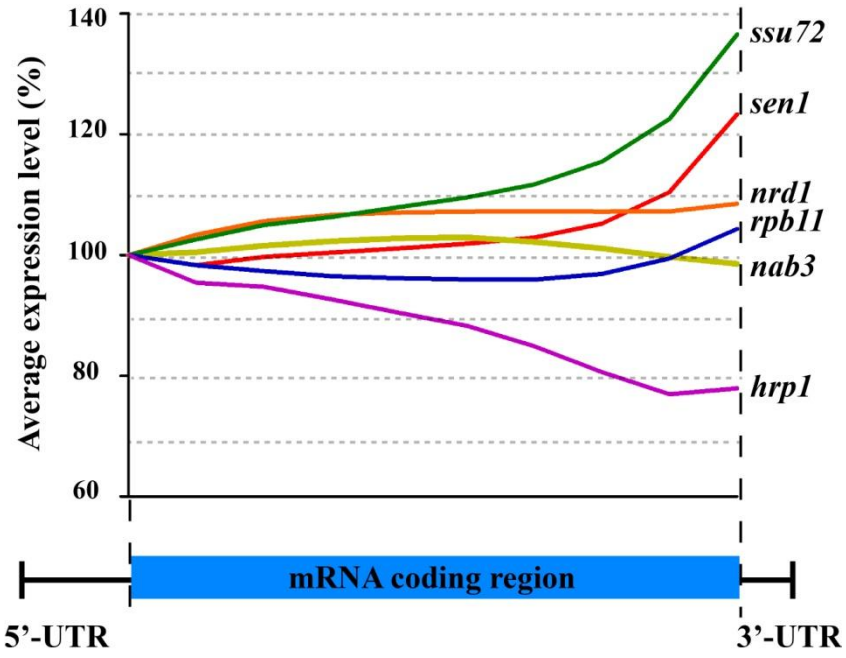
differential fold change value of each segment, which represents the transcript level of this segment in the mutant relative to wild-type. By setting the value of the segment next to the 5'-UTR to 100%, I normalized the values of the remaining nine segments for every protein-coding gene, and then calculated the average values of each of the 10 segments.

As shown in Figure 4.3, the transcript level changes throughout protein-coding genes are always less than 10% in the Nrd1, Nab3 and Rpb11 mutants, indicating they do not have a strong graduated effect on the transcript level along mRNA genes in general. In the Sen1 mutant, there is a significant increase of transcript level at the 3' end, in concert with the previous cross-linking result showing Sen1-enrichment at the 3' end of pre-mRNA transcripts (CREAMER *et al.* 2011). This finding supports the proposed function of Sen1 in transcription termination of some protein-coding genes. Alternatively, this increase may represent antisense transcripts that arise upon dysfunction of Sen1 and accumulate around the 3' ends of protein-coding genes.

I observed a gradual decrease of transcript level along the length of mRNA genes in the Hrp1 mutant. This finding is consistent with the observation that Hrp1 depletion decreases the transcript level of some protein-coding genes *in vivo*, which Tuck and Tollervery attributed to an essential role of Hrp1 in mRNA synthesis (TUCK and TOLLERVEY 2013). However, the decrease caused by the mutation is inconsistent with the function of Hrp1 in mRNA attenuation (KUEHNER and BROW 2008). Since Hrp1 associates with not only the 3'-end poly(A) signals but also throughout the promoters and transcribed regions of mRNA genes (KOMARNITSKY *et al.* 2000), the effect of decreasing transcript level caused by the Hrp1 substitution suggests a role of Hrp1 in maintaining the transcription processivity of Pol II on mRNA genes. Nevertheless, the mechanism underlying the decreasing effect is to be identified.

In contrast, the transcript level is gradually increased along the full-length of mRNA

FIGURE 4.3 Metagene profile shows effects of the six mutants on expression levels across the coding region of mRNA genes. The coding region of every mRNA gene was divided into 10 segments evenly, and the average transcript level of each fragment across all mRNA genes was calculated. The values of all fragments are normalized to that of the fragment next to the 5'-UTR, which is set to 100%. The y-axis shows the average expression level (in percentage) of each fragment. The x axis is aligned to the schematic diagram of an mRNA coding gene. The expression profiles in the six mutants are shown in different colors.


140
120
100
80
60
Average expression level (%)
ssu72
sen1
nrd1
rpb11
nab3
hrp1
mRNA coding region
5'-UTR
3'-UTR

genes in the Ssu72 mutant. Previously, Zhang *et al*. observed Pol II occupancy is lower during early elongation of protein-coding genes in another phosphatase-defective Ssu72 mutant (ZHANG *et al.* 2012). The gradually increasing distribution of Pol II along the length of protein-coding genes they observed resembles the pattern of transcript level change identified in our Ssu72 mutant. They proposed that inactivation of Ssu72 prevents the dephosphorylation of the CTD, which impedes Pol II re-assembly into the promoter-bound preinitiation complex and thereby decreases the reinitiation of the next round of transcription. However, what causes the uneven increase of transcript level across the mRNA coding regions in the Ssu72 mutant is unclear. As Ssu72 is a component of the cleavage/polyadenylation factor (CPF) and required for pre-mRNA cleavage (HE *et al.* 2003), perhaps dysfunction of Ssu72 counteracts cleavage of the nascent transcript at the internal poly(A) sites which decreases transcript level toward the 3' end. Therefore, the transcript level in the mutant exhibits a gradual increase along the direction of transcription relative to that in wild-type.

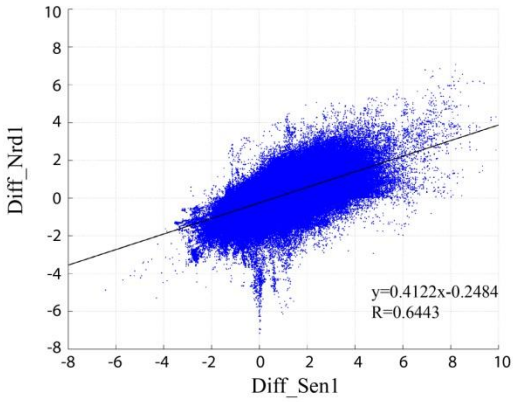
4.4 The transcripts affected by different mutants only partially overlap

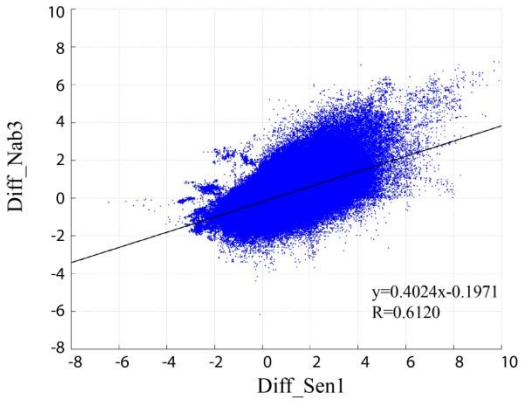
While read-through of some Sen1-dependent terminators was induced by most or all of the substitutions, other terminators were responsive to only one or a few of them. Furthermore, some changes in transcript accumulation induced by the mutations were not obviously related to terminator read-through, such as the transcripts that are strongly up-regulated in the Rpb11 or Hrp1 mutant but not increased in the Sen1, Nrd1 or Nab3 (components of the core termination complex) mutants. To identify the functional convergence among the six proteins across the genome, I calculated the degree to which changes in transcript level over each probe (5 base-pair interval) are correlated for each pairwise combination of mutants (Table 4.2 and Figure 4.4).

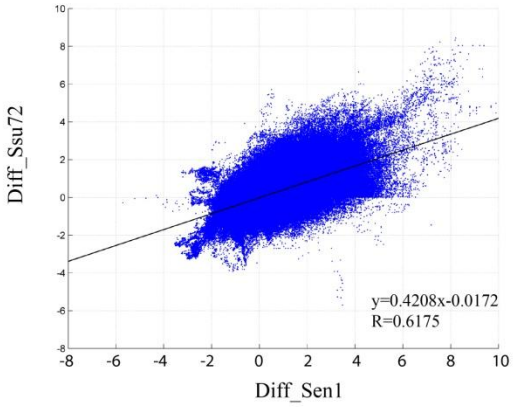
TABLE 4.2 Summary of the Pearson's correlation coefficients of each pairwise combination of mutants

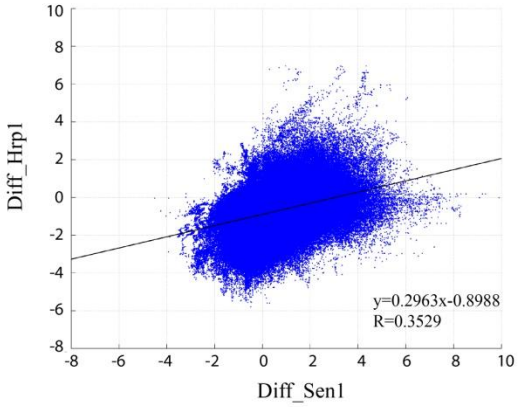
Group	Mutants		Pearson's correlation
I	*nrd1*	*nab3*	0.66
	sen1	*nrd1*	0.64
	ssu72	*nrd1*	0.62
	sen1	*ssu72*	0.62
	sen1	*nab3*	0.61
II	*ssu72*	*nab3*	0.52
III	*nrd1*	*hrp1*	0.37
	nab3	*hrp1*	0.37
	sen1	*hrp1*	0.35
IV	*sen1*	*rpb11*	0.34
V	*ssu72*	*hrp1*	0.24
	ssu72	*rpb11*	0.21
VI	*nab3*	*rpb11*	0.17
	nrd1	*rpb11*	0.12
	hrp1	*rpb11*	0.08

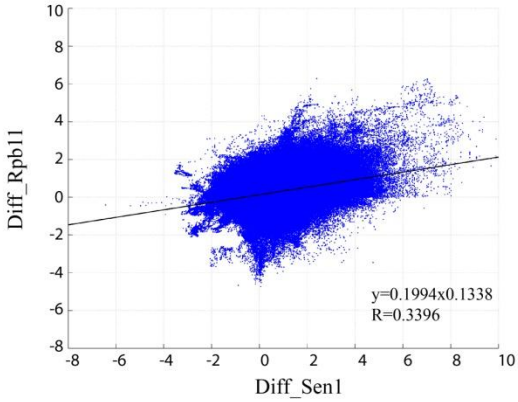
FIGURE 4.4 Scatter plots of fold change in transcript level relative to the wild-type. The data points in the scatter plots represent the pairwise differential transcript levels (log2 scale) in two mutants in ~5 base-pair increment across the genome. The x- and y-axes show the differential transcript levels. Fifteen scatter plots cover all the possible pairwise combinations. The equation of the linear regression line and the Pearson's correlation coefficient (R) are indicated on every scatter plot. A smoothing window of 101 bp was used for these plots.

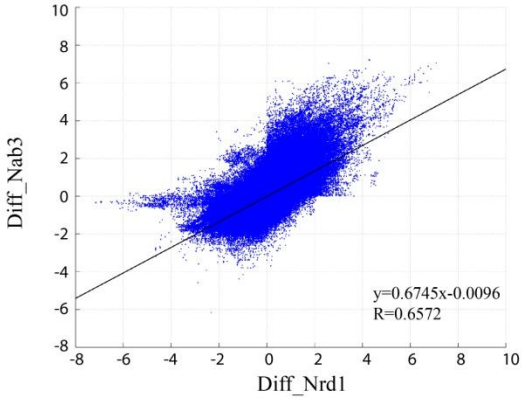

Diff_Nrd1
Diff_Sen1
y=0.4122x-0.2484
R=0.6443

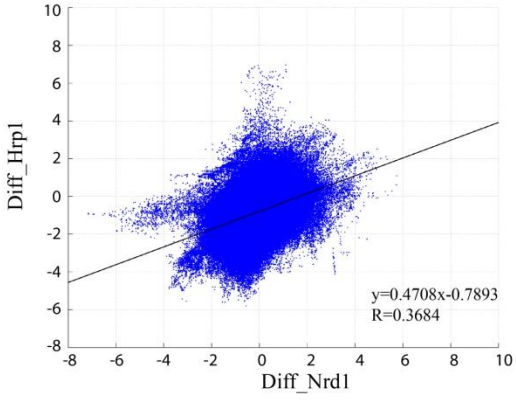

Diff_Nab3
Diff_Sen1
y=0.4024x-0.1971
R=0.6120

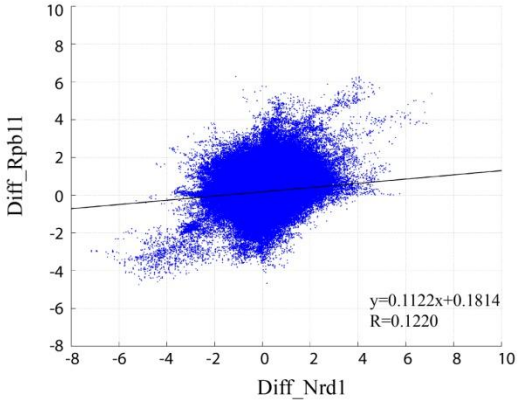

Diff_Ssu72
Diff_Sen1
y=0.4208x-0.0172
R=0.6175

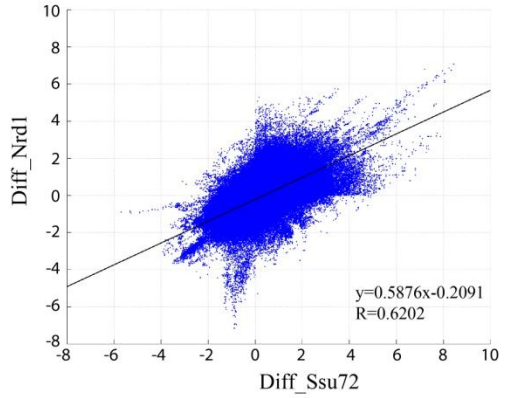

Diff_Hrp1
Diff_Sen1
y=0.2963x-0.8988
R=0.3529

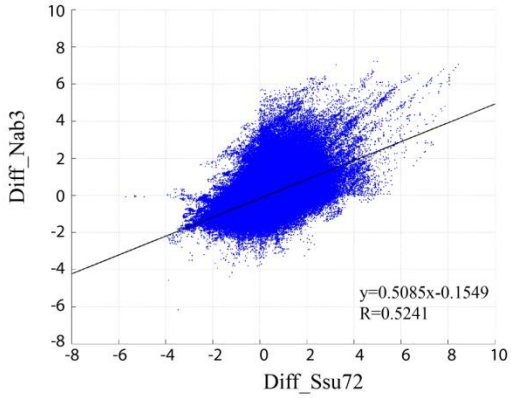

Diff_Rpb11
Diff_Sen1
y=0.1994x0.1338
R=0.3396

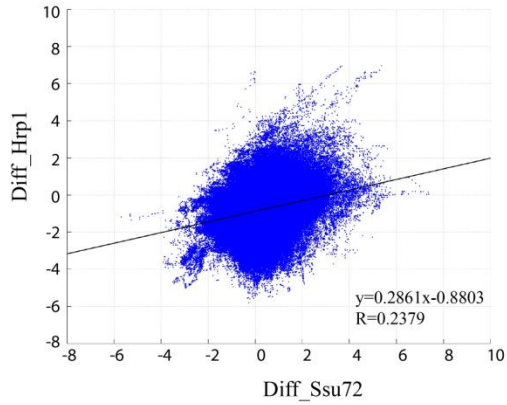

Diff_Nab3
Diff_Nrd1
y=0.6745x-0.0096
R=0.6572

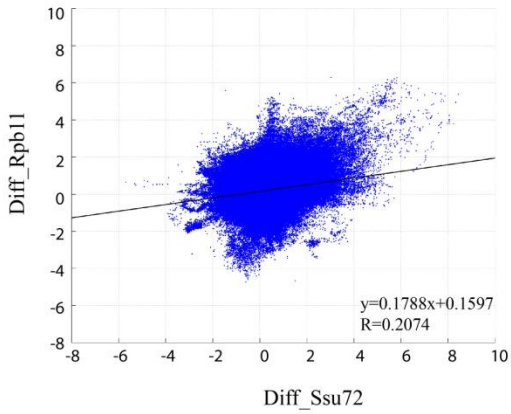

Diff_Hrp1
Diff_Nrd1
y=0.4708x-0.7893
R=0.3684

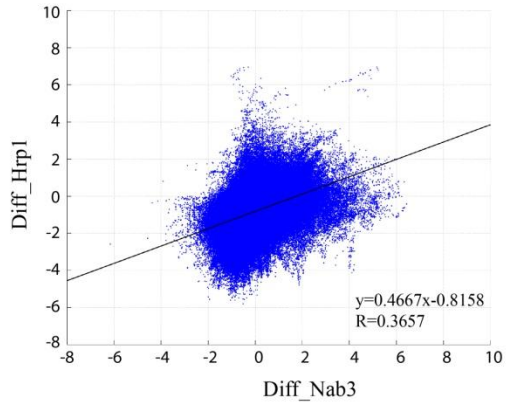

Diff_Rpb11
Diff_Nrd1
y=0.1122x+0.1814
R=0.1220

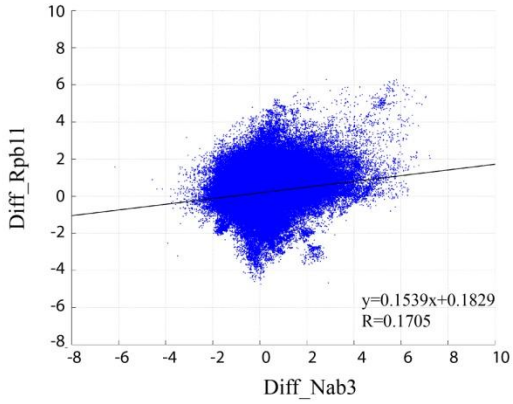

Diff_Nrd1
Diff_Ssu72
y=0.5876x-0.2091
R=0.6202

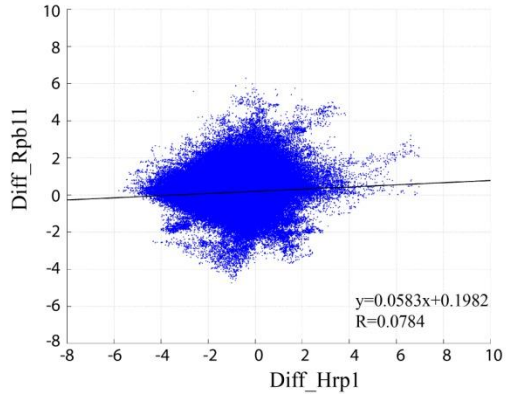

Diff_Nab3
Diff_Ssu72
y=0.5085x-0.1549
R=0.5241


Diff_Hrp1
Diff_Ssu72
y=0.2861x-0.8803
R=0.2379


Diff_Rpb11
Diff_Ssu72
y=0.1788x+0.1597
R=0.2074


Diff_Hrp1
Diff_Nab3
y=0.4667x-0.8158
R=0.3657


Diff_Rpb11
Diff_Nab3
y=0.1539x+0.1829
R=0.1705


Diff_Rpb11
Diff_Hrp1
y=0.0583x+0.1982
R=0.0784

According to the calculated Pearson's correlation coefficients, all the mutants are positively correlated yet to different degrees, and I categorized their pairwise correlations into six groups based on the correlation degrees and the constitutive factors.

Group I includes all the combinations among Sen1, Nrd1 and Nab3, the three known components of the core termination complex. The highest correlation is between the Nrd1 and Nab3 mutants, which has a pairwise correlation coefficient of 0.66, consistent with the evidence that they function in the form of heterodimer. The Sen1 mutant highly correlates with the Nrd1 mutant and the Nab3 mutant, with the pairwise correlation coefficients 0.64 and 0.61, which also agrees with the model that the heterodimer of Nrd1 and Nab3 recruits Sen1 for termination. The pairs of Sen1-Ssu72 and Nrd1-Ssu72 also fall into this group, and both have a pairwise correlation coefficient of 0.62. This correlation agrees with the regulatory function of Ssu72's phosphatase activity on the Pol II CTD to mediate the recruitment of the Sen1 machinery, and the interactions between the CTD and Sen1 or Nrd1 (STEINMETZ and BROW 1998; CONRAD *et al.* 2000; VASILJEVA *et al.* 2008; CHINCHILLA *et al.* 2012).

The pair of Nab3 and Ssu72 makes Group II, with a pairwise correlation coefficient of 0.52. No direct interaction has been identified between Nab3 and the Pol II CTD, which may explain the lower correlation between Ssu72 and Nab3 than that between Ssu72 and Nrd1, although Nrd1 and Nab3 are proposed to have similar function during Sen1-dependent termination. The decreased correlation suggests that, in some cases, the CTD-regulated Sen1-dependent termination does not require Nab3 or its RNA binding activity that is affected by the substitution.

Group III includes the combinations between Hrp1 and all three components of the core termination complex (correlation coefficients 0.35-0.37). This moderate correlation suggests that

Hrp1 has a broad influence on Sen1-dependent termination. A chromatin immunoprecipitation experiment showed that Hrp1 localizes to snoRNA genes (KIM *et al.* 2006). Tuck and Tollervey found Hrp1 is distributed over CUTs, and enriched at promoter-proximal regions of mRNAs, suggesting Hrp1 could participate in the nuclear turnover of CUTs (TUCK and TOLLERVEY 2013). In agreement with this hypothesis, Kuehner and Brow confirmed that the Hrp1 substitution causes read-through of the *HRP1* and *NRD1* attenuators (KUEHNER and BROW 2008).

Group IV is the pair of Sen1 and Rpb11. The correlation coefficient between these two mutants (0.34) is higher than that between the Rpb11 mutant and the mutants of Nrd1 or Nab3 (0.12 or 0.17), suggesting that there is certain function of Sen1 and Pol II that is independent of Nrd1 and Nab3. Given the position of Rpb11 in Pol II (STEINMETZ *et al.* 2006a), such function may rely on the interaction between Sen1 and the Rpb11 subunit of Pol II. In fact, Sen1 may recognize and bind to specific RNA sequences by itself. When purified from *E. coli*, the recombinant Sen1 helicase domain was associated with short RNAs that are enriched for the tri-nucleotide repeat $(CAN)_4$ (S. Martin-Tumasz and D. Brow, unpublished data). The Nrd1- and Nab3-indepedent RNA-binding ability of Sen1 may allow it to interact with Rpb11.

Group V includes the pair of Ssu72 and Hrp1 mutants (correlation coefficient 0.24), followed by the Ssu72 and Rpb11 combination (correlation coefficient 0.21). Because Ssu72 and Hrp1 are both associated with the cleavage and polyadenylation machinery (KESSLER *et al.* 1997; HE *et al.* 2003), the correlation between them probably reflects their involvement in mRNA 3' end processing. On the other hand, the role of the CTD may contribute to the correlation between Ssu72 and Rpb11. By regulating the phosphorylation state of the CTD (KRISHNAMURTHY *et al.* 2004), Ssu72 may be involved in the recruitment of Rpb11-interacting transcription factors.

Group VI is between the Rpb11 mutant and the mutants in Nab3, Nrd1 or Hrp1. The

correlation coefficients are low (0.17, 0.12 and 0.08), suggesting the Rpb11 mutation affects genes that are often not sensitive to the mutations of the other three, and vice versa.

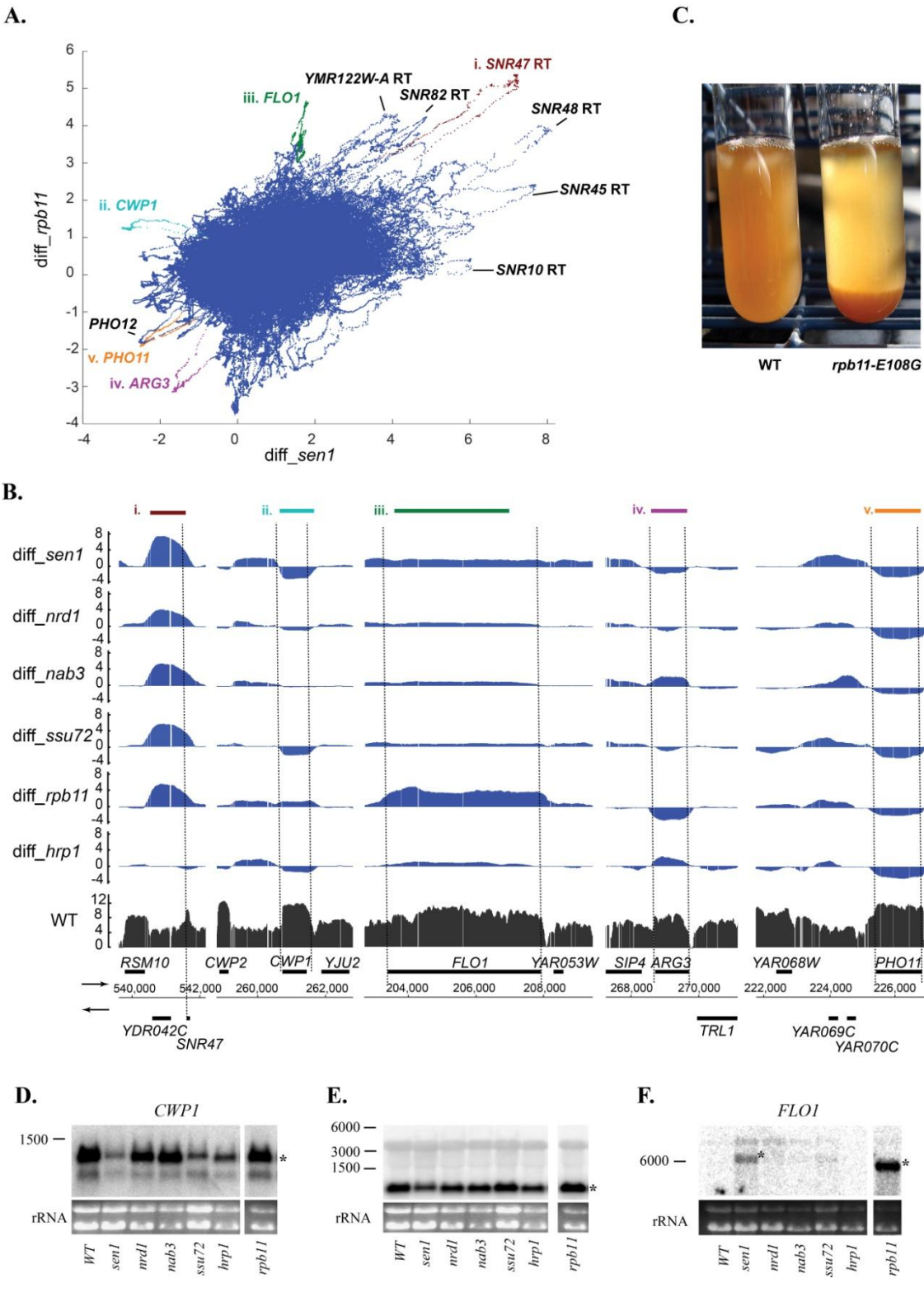
4.5 A discrete loop pattern from the pairwise scatter plot facilitates transcription unit identification

In addition to correlation coefficients, pairwise scatter plots with a large smoothing window (500 bp) show an interesting feature, namely discrete loops that extend from the mass of data points. These loops represent individual transcription units whose expression levels are most significantly changed by one or both mutations.

In the case of the Sen1 vs. Rpb11 scatter plot (Figure 4.5A), most of the loops corresponding to high differential expression in both mutants are snoRNA gene terminator read-through transcripts. However, the effect of the Rpb11 substitution on read-through of these terminators varies from high in the case of *SNR47* to low in the case of *SNR10* (Figures 4.5A and 4.5B.i). Transcripts that are strongly decreased in the Sen1 mutant but not the Rpb11 mutant include *CWP1* (Figures 4.5A and 4.5B.ii). By Northern blot analysis using a strand-specific oligo probe, I confirmed the *CWP1* levels are decreased in the Sen1, Ssu72 and Hrp1 mutants (Figure 4.5D), which is consistent with the transcriptome data. What causes the decrease of *CWP1* level? Upon dysfunction of Sen1, transcripts from terminator read-through of *CWP2* or initiation from the *CWP1-CWP2* intergenic region may decrease the *CWP1* level by promoter occlusion. However, my Northern blot analysis using strand-specific oligo probes did not detect either *CWP2* read-through or increased promiscuous transcript from the intergenic region (Figure 4.5E).

There are genes whose transcripts accumulated to a higher level in the Rpb11 mutant than the Sen1 mutant, such as *FLO1* (Figures 4.5A and 4.5B.iii), which promotes flocculation. The increased expression of the *FLO1* gene is accompanied by flocculation in the Rpb11 mutant

FIGURE 4.5 Discrete loops in the pairwise scatter plots reveal transcription units that are significantly affected by the substitutions. A) Scatter plot of log2 fold change in transcript level relative to the wild-type over every ~5 base-pair interval in the Sen1 mutant (x-axis) vs. the Rpb11 mutant (y-axis) using a 500 bp smoothing window. Loops are labeled according to the associated transcription units, and the highlighted color loops are presented in 4.4B. **B)** Changes in transcript level of the highlighted transcription units in all six mutants. The colored bars indicate the genome location of the corresponding loops in Panel A. The blue traces represent the differential transcript levels in the indicated mutants, and the black traces represent the absolute transcript level in wild-type. The annotated transcription units are indicated by the black bars. Transcription units on the Watson stand are above the chromosomal line and on the Crick strand are below. **C)** The *rpb11-E108G* mutant (right) exhibits a strong flocculation phenotype compared to the wild-type (left) in YEPD medium. **D**), **E**) and **F**) Total RNA from haploid strains harboring wild-type alleles (WT) or the indicated mutant alleles was resolved on denaturing agarose gels and detected by the *CWP1*, *CWP2* and *FLO1* oligo probes. The numbers to the left of the blots represent the length in nucleotides of RNA marker bands. The positions of the *CWP1* and *FLO1* transcript bands are indicated by the asterisk. Ethidium bromide stain of the portion of the filter containing the large and small ribosomal RNAs is shown below the blots.


A.
C.
B.
D.
E.
F.
i. SNR47 RT
YMR122W-A RT
iii. FLO1
SNR82 RT
SNR48 RT
SNR45 RT
ii. CWP1
SNR10 RT
PHO12
v. PHO11
iv. ARG3
diff_rpb11
diff_sen1
WT
rpb11-E108G
diff_sen1
diff_nrd1
diff_nab3
diff_ssu72
diff_rpb11
diff_hrp1
WT
RSM10
CWP2
CWP1
YJU2
FLO1
YAR053W
SIP4
ARG3
YAR068W
PHO11
YDR042C
SNR47
TRL1
YAR069C
YAR070C
CWP1
rRNA
FLO1
WT
sen1
nrd1
nab3
ssu72
hrp1
rpb11

(Figures 4.5C). The increased FLO1 levels in the Sen1 and Rpb11 mutants were confirmed by Northern blot analysis using a strand-specific oligo probe (Figure 4.5F). The Rpb11 mutant also decreased the expression of every gene in the arginine synthesis pathway, including *ARG3* (Figures 4.5A and 4.5B.iv). The two repressible acid phosphatase paralogs, *PHO11* and *PHO12*, both exhibited decreased mRNA levels in response to all six mutations (Figures 4.5A and 4.5B.v).

4.6 sn/snoRNA genes exhibit differential sensitivity to substitutions in the six Sen1-dependent termination-related proteins

A primary target of Sen1-dependent termination is sn/snoRNA genes. To see how the read-through substitutions in the six proteins affect termination of these genes, I quantified sn/snoRNA gene transcription termination efficiency in wild-type and mutant strains. The percentage read-through ratio is calculated by dividing the average transcript level in a 50 base-pair window downstream of the gene by the average level of the entire sn/snoRNA itself and multiplying by 100%. The percentage differential read-through ratio between the mutant and its wild-type control (read-through ratio in mutant subtracted by read-through ratio in WT) was calculated for each gene to quantify the read-through effect of this mutation on every sn/snoRNA gene (Table 4.3). In principle, a read-through ratio of 100% indicates no termination, while 0% differential read-through ratio means no termination defect. I excluded from the analysis sn/snoRNA genes with closely adjacent or overlapping transcription units at their 3' ends, leaving 48 Pol II-transcribed genes (or poly-cistronic gene clusters) with sufficient signal quality for the read-through ratio analysis. In some cases, the differential read-through ratio is above 100%, which may be due to antisense transcripts accumulating near the 3' ends of the sn/snoRNA genes in the mutants.

TABLE 4.3 Summary of the differential read-through ratio (%) of sn/snoRNA genes in the six mutants.

SNR	*sen1* [a]	*nrd1* [a]	*nab3* [a]	*ssu72* [a]	*rpb11* [a]	*hrp1* [a]
SNR33	**342.00**	**91.84**	**55.18**	**253.01**	**143.24**	**55.28**
SNR48	**236.27**	**94.56**	**178.67**	**89.58**	**39.32**	**18.01**
SNR161	**231.51**	**223.36**	**212.37**	**148.84**	**35.16**	**112.28**
SNR31	**175.48**	**24.73**	**18.91**	**29.31**	-2.15	**23.18**
SNR79	**162.34**	7.37	**62.63**	**39.43**	**12.71**	9.31
SNR80	**161.12**	**60.90**	**81.84**	**115.87**	**36.91**	**140.46**
SNR51	**137.87**	**16.99**	-2.67	**141.70**	**48.38**	**34.35**
SNR47	**131.89**	**28.24**	**86.95**	**58.17**	**36.07**	**11.44**
SNR82	**124.83**	**41.30**	**14.26**	**45.11**	**57.60**	**87.55**
SNR37	**109.78**	**13.62**	**20.84**	9.08	0.80	1.82
SNR61	**106.65**	-5.08	**14.87**	**42.45**	4.78	**40.81**
SNR50	**106.12**	9.72	6.41	**38.01**	6.17	7.21
SNR85	**98.99**	-2.51	1.45	**10.06**	-0.24	-0.84
SNR87	**86.90**	-16.44	**37.79**	**71.02**	**21.89**	-49.28
SNR72	**85.43**	-10.29	**12.56**	**39.28**	**60.65**	-9.02
SNR34	**82.83**	**33.41**	**43.59**	7.75	4.89	2.64
SNR64	**74.69**	**32.90**	**21.55**	**34.52**	7.97	2.18
SNR11	**72.94**	**34.20**	4.61	**32.55**	8.10	9.64
SNR42	**69.46**	**38.83**	**35.94**	**58.68**	-14.24	**62.41**
SNR45	**68.59**	6.12	**17.20**	2.31	4.43	-3.91
SNR81	**67.01**	6.63	**13.98**	**38.26**	**10.53**	-0.02
SNR3	**66.68**	**16.01**	**59.17**	9.63	5.76	-3.95
SNR13	**66.60**	**44.05**	**32.55**	**33.98**	5.87	1.45
SNR189	**62.67**	0.33	7.24	**27.15**	9.35	-0.38
SNR39B	**60.33**	3.97	0.81	9.75	1.46	-0.49
SNR35	**57.01**	**15.59**	2.20	**11.27**	-1.08	7.51
SNR49	**46.84**	**42.25**	**10.06**	-116.76	-85.00	**209.31**
SNR69	**45.31**	**26.09**	**31.69**	**22.07**	6.40	**48.53**
SNR128	**39.21**	2.04	**10.00**	4.80	2.34	-0.22
SNR32	**39.01**	5.49	5.85	2.98	5.86	1.69
SNR71	**38.21**	**25.25**	**45.31**	**37.93**	**19.15**	**31.88**
SNR14(U4)	**36.37**	**48.92**	**21.03**	**25.58**	-10.66	-14.27
SNR8	**35.80**	0.05	0.75	3.99	2.71	-0.06
SNR5	**34.22**	**14.96**	**13.92**	**19.12**	8.45	3.80
SNR10	**28.48**	**10.96**	-3.41	4.74	3.24	-7.63
SNR4	**23.90**	**10.03**	**17.31**	**15.95**	4.12	**10.58**
SNR30	**23.65**	5.93	2.96	0.84	1.26	0.49
SNR46	**15.57**	4.14	3.59	2.50	2.06	-1.26
SNR84	**15.29**	**10.56**	**20.96**	**11.10**	-4.61	**19.95**
SNR86	**12.17**	2.31	-1.18	4.47	-1.97	-2.03
SNR63	8.02	**10.88**	-3.81	6.07	-0.11	-3.21
SNR17b	7.14	9.92	-1.49	4.92	-1.82	-0.25
SNR7 (U5)	5.85	1.82	**14.10**	0.50	-0.22	-0.97
SNR19(U1)	3.50	1.72	1.15	3.02	0.45	0.10
SNR17a	3.41	1.43	1.01	2.02	0.70	0.27

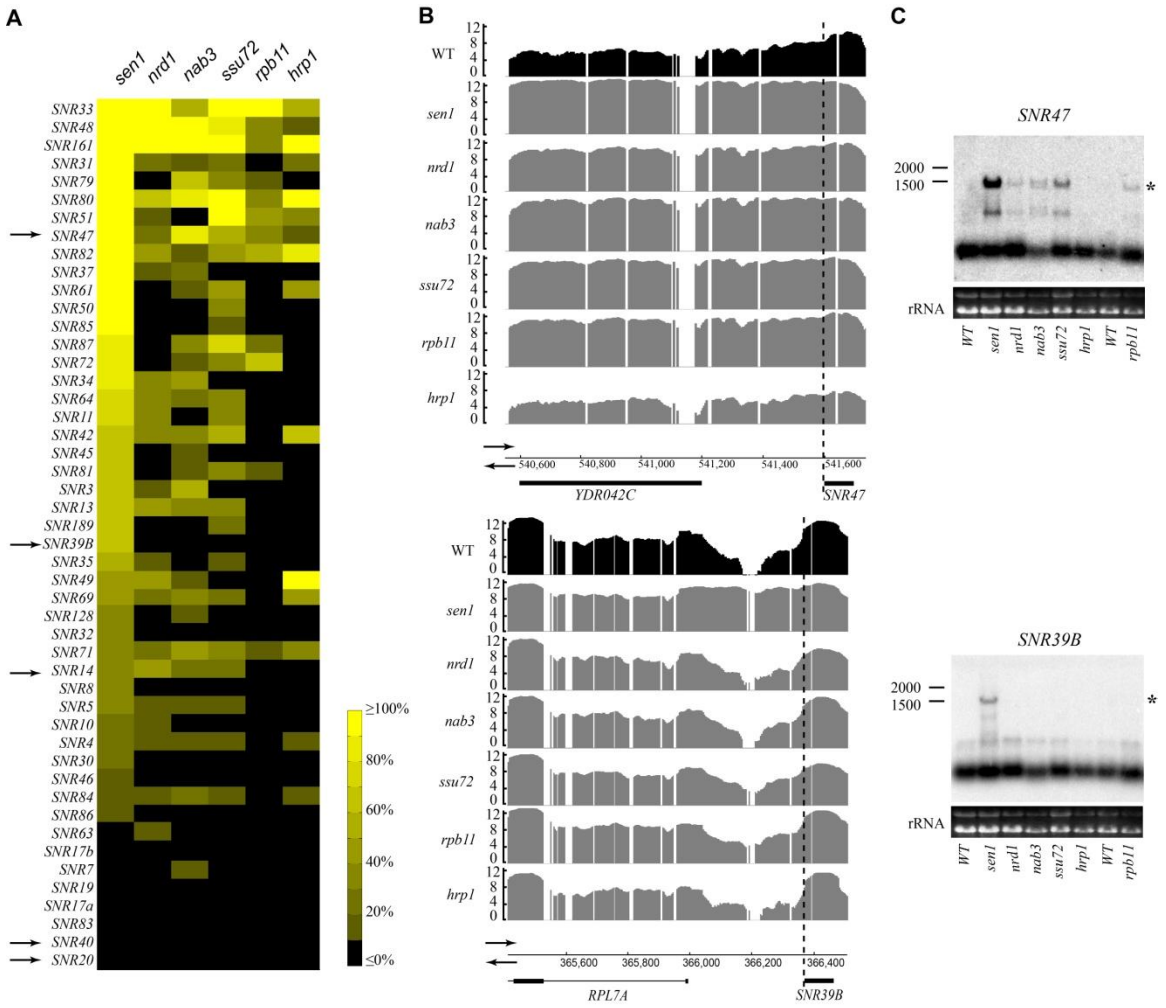
SNR83	3.30	1.43	-0.70	-0.08	-0.23	-2.47
SNR40	2.89	0.08	-0.17	0.64	4.40	0.29
SNR20(U2)	-0.24	-0.42	-1.24	-0.21	-2.80	-3.62

[a]: Differential read-through ratios more than 10% are considered as termination defects and highlighted in bold font.

To highlight significant termination defects, I consider over 10% differential read-through ratio bone fide termination defect. According to this criterion, 42 Pol II-transcribed sn/snoRNA genes showed termination defects in at least one mutant (see Figure 4.6A, where yellow squares indicate termination defects). Efficient termination of sn/snoRNA genes exhibited differential factor requirement. For instance, termination of *SNR47* is read-through in all the mutants, but to varying degrees (Figure 4.6B, top). Indeed, extended transcripts of *SNR47* were detected by Northern blot in all the mutants except Hrp1 (Figure 4.6C, top), where the false positive result may be caused by decreased expression of the snoRNA itself. In contrast, both the transcriptome data and the Northern blot analysis showed that termination of *SNR39B* is only defective in the Sen1 mutant (Figures 4.6B and C, bottom). Notably, on both *SNR47* and *SNR39B* blots, intermediate bands with sizes between the mature snoRNAs and the extended products terminated at the downstream mRNA genes' poly(A) sites were observed. As there are intergenic poly(A) sites between the snoRNA genes and their downstream mRNA genes (OZSOLAK *et al.* 2010), the intermediate products are likely to represent the snoRNA extended transcripts that pass the Sen1 terminators but stop at the intergenic poly(A) sites.

Among the six mutants, the Sen1 mutant showed the broadest influence, causing termination defects on 40 Pol II-transcribed sn/snoRNA genes (Figure 4.6A and Table 4.3), including both H/ACA box snoRNAs and C/D box snoRNAs. This finding argues against the reported target preference of Sen1 for H/ACA box snoRNA genes over C/D box snoRNA genes based on the Sen1 occupancy (JAMONNAK *et al.* 2011). The only snRNA gene that exhibited a termination defect in the Sen1 mutant is *SNR14*. Consistent with this finding, Sen1-dependent terminator elements have been identified downstream of this snRNA gene previously, about 380 base pairs away from its transcription start site (STEINMETZ *et al.* 2001).

FIGURE 4.6 Effects of the six mutations on transcription termination of sn/snoRNA genes. **A)** A heat map summarizing the differential read-through ratio of indicated sn/snoRNA genes in the six mutants ranked from high to low according to the differential read-through value in the Sen1 mutant. The genes that are specifically mentioned in Section 4.7 are indicated by arrows. **B)** Termination read-through of *SNR47* and *SNR39B* in all six mutants. The black traces represent the absolute transcript level in wild-type, and the grey traces represent the absolute transcript level in the mutants (in log2 scale). The annotated genes are indicated by the black bars. Genes on the Watson stand are above the chromosomal line and on the Crick strand are below. **C)** Northern blot analyses of *SNR47* and *SNR39B* RNAs in the mutants. Total RNA from the indicated strains was resolved on denaturing agarose gels. The numbers to the left of the blots represent the length in nucleotides of RNA markers. The positions of the snoRNA read-through transcripts that are terminated at the downstream mRNA gene poly(A) sites are indicated by asterisks. Ethidium bromide stain of the filter showing the large and small ribosomal RNAs is shown below the blots.


A
sen1
nrd1
nab3
ssu72
rpb11
hrp1
SNR33
SNR48
SNR161
SNR31
SNR79
SNR80
SNR51
SNR47
SNR82
SNR37
SNR61
SNR50
SNR85
SNR87
SNR72
SNR34
SNR64
SNR11
SNR42
SNR45
SNR81
SNR3
SNR13
SNR189
SNR39B
SNR35
SNR49
SNR69
SNR128
SNR32
SNR71
SNR14
SNR8
SNR5
SNR10
SNR4
SNR30
SNR46
SNR84
SNR86
SNR63
SNR17b
SNR7
SNR19
SNR17a
SNR83
SNR40
SNR20
≥100%
80%
60%
40%
20%
≤0%
B
WT
sen1
nrd1
nab3
ssu72
rpb11
hrp1
12
8
4
0
540,600
540,800
541,000
541,200
541,400
541,600
YDR042C
SNR47
365,600
365,800
366,000
366,200
366,400
RPL7A
SNR39B
C
SNR47
2000
1500
*
rRNA
WT
sen1
nrd1
nab3
ssu72
hrp1
WT
rpb11
SNR39B

Most of the genes that do not exhibit over 10% read-through in the Sen1 mutant are relatively long (>400 bps), such as *SNR20*, whose mature transcript has a length of 1175 nucleotides (Figure 4.6A and Table 4.3). In agreement with this finding, Steinmetz *et al*. reported that *SNR20* did not exhibit terminator read-through in the Sen1 mutant (STEINMETZ *et al.* 2006b). This observation agrees with the notion that the Sen1 pathway preferentially terminates transcription of short genes (usually <450 bps) (STEINMETZ *et al.* 2006b; GUDIPATI *et al.* 2008; JENKS *et al.* 2008; VASILJEVA *et al.* 2008). However, *SNR40*, whose mature transcript is 97 nucleotides long, can be efficiently terminated in the Sen1 mutant. According to the location of *SNR40*'s downstream poly(A) sites, the primary transcript of this snoRNA gene is unlikely to be over 400 bp long (OZSOLAK *et al.* 2010), within the range of Sen1-dependent termination. According to our Pol II ChIP-chip data, the Pol II peak on *SNR40* apparently shifts downstream in the Sen1 mutant (STEINMETZ et al. 2006b), indicating termination read-through. Perhaps there are endoribonuclease Rnt1 sites in the *SNR40* read-through transcript (ABOU ELELA *et al.* 1996; ALLMANG *et al.* 1999). Cleavage by Rnt1 and degradation of the read-through product prevent its accumulation. Therefore, I did not observe a significant read-through effect of this gene in the Sen1 mutant.

Transcription termination of sn/snoRNA genes is less often affected by the substitutions in the other five proteins. The Nrd1, Nab3 and Ssu72 mutants cause termination defects on about two thirds as many sn/snoRNA genes as Sen1, on 25, 29 and 27 genes, respectively. All these genes also exhibited termination defects in the Sen1 mutant except *SNR63* and *SNR7*, and the differential read-through ratios of these two genes are barely above the 10% cutoff in the Nrd1 and Nab3 mutants, respectively. Fewer sn/snoRNA genes show termination defects in the Rpb11 and Hrp1 mutants, 12 and 16 respectively. They are all identified as termination-defective in the

Sen1 mutant.

In summary, Sen1-dependent termination appears to target short, Pol II-transcribed sn/snoRNA genes, and the activity of Sen1 is indispensible for efficient termination of over 80% of them, while dysfunction of the other five proteins each affects transcription termination of only a subset of the Sen1's target genes.

4.7 Nrd1 and Nab3 are not functionally redundant for transcription termination of selected snoRNA genes

As revealed by the earlier analyses in this chapter, substitutions in the RRMs of Nrd1 and Nab3 appear to only affect transcription termination of a subset of Sen1's target genes. For example, the transcriptome data show that snoRNAs can be sensitive to both Nrd1 and Nab3 mutants (e.g., *SNR47*), Nab3 alone (e.g., *SNR79*), Nrd1 alone (e.g., *SNR11*) and neither (e.g., *SNR39B*) (Figure 4.6A). This finding suggests the RNA binding abilities of Nrd1 and Nab3 are not always required for Sen1-dependent termination. Alternatively, Nrd1 and Nab3 may be functionally interchangeable for terminating transcription, as high copy *NAB3* was found to suppress an *nrd1* mutant allele, and such functional complementation may further be facilitated by forming the higher order Nrd1-Nab3 oligomer complex (CONRAD *et al.* 2000; LOYA *et al.* 2013a), so the effects of Nrd1 and Nab3 on termination can be underestimated by analyzing them individually.

To assess redundant effects of Nrd1and Nab3 on transcription termination, I need a haploid strain that bears both *nrd1-5* and *nab3-11* alleles. Previous efforts to create viable Nrd1-Nab3 double mutants were not successful, as such double mutants are either lethal or very slow growing at 30 ℃ (LOYA *et al.* 2013a). Because the *nab3-11* mutant is lethal on plates at 30 ℃, but viable at 23 ℃, I transformed the pRS313-nab3-11 plasmid into a *nrd1-5* mutant strain in

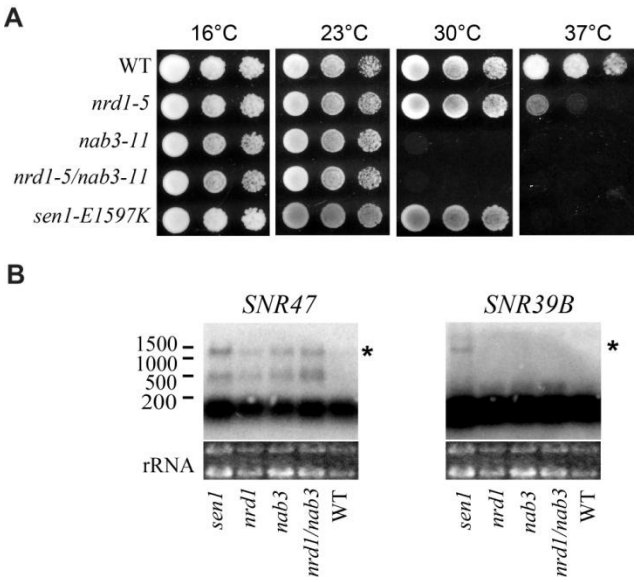
which chromosomal *NAB3* is replaced with pRS316-NAB3, and performed the 5-FOA plasmid shuffle experiment at 23 ℃. I successfully obtained the double mutant (XCY377). This strain shows no growth defect at 16 ℃ and 23 ℃, but is lethal at 30 ℃ and 37 ℃ on the YEPD plate (Figure 4.7A). Its growth phenotypes at these temperatures resemble that of the *nab3-11* mutant. As the substitutions are in the RRMs of Nrd1 and Nab3, they are expected to disrupt the RNA binding abilities of Nrd1 and Nab3simultaneously.

Using Northern blot assays, I investigated transcripts of *SNR47* and *SNR39B* in the wild-type, *sen1-E1597K*, *nrd1-5*, *nab3-11* and the double mutant XCY377 strains. In the case of *SNR47* (Figures 4.6C, top and 4.7B, left), termination read-through was detected in both Nrd1 and Nab3 mutants. In XCY377, I observed increased accumulation of extended transcripts representing read-through products compared to the Nrd1 and Nab3 single mutants, similar to the read-through level in the Sen1 mutant. These results indicate efficient termination of *SNR47* required the RNA binding abilities of both Nrd1 and Nab3, but Nrd1 and Nab3 do not appear to be functionally interchangeable on this snoRNA gene. For *SNR39B* (Figures 4.6C, bottom and 4.7B, right), the extended transcript representing *SNR39B* read-through products was only detected in the Sen1 mutant but not the others. Therefore, the absence of read-through in the presence of Nrd1 and Nab3 substitutions is not due to functional redundancy of these two factors.

4.8 Sen1-dependent termination represses meiotic gene expression in vegetative cells

Although Sen1 broadly affects gene expression by repressing promiscuous transcription, the Sen1-dependent transcription termination pathway represses expression of specific genes by mechanisms such as attenuation. Based on our transcriptome data, there are 769 mRNA genes showing at least a two-fold net increase in the *sen1-E1597K* mutant, although some of these are due to read-through of adjacent snoRNA genes or unannotated transcription units. According to

FIGURE 4.7 Effects of the Nrd1/Nab3 double mutant on transcription termination of *SNR47* and *SNR39B* genes. A) Growth phenotypes at 16, 23, 30 and 37℃ of the wild-type, Sen1, Nrd1 and Nab3 single mutants, and the Nrd1/Nab3 double mutant. **B)** Northern blot analyses of *SNR47* and *SNR39B* RNAs in the indicated mutants. Total RNA from the indicated strains was resolved on denaturing agarose gels. The numbers to the left of the blots represent the length in nucleotides of RNA marker bands. The positions of the snoRNA read-through transcripts that are terminated at the downstream mRNA gene poly(A) sites are indicated by asterisks. Ethidium bromide stain of the filter showing the large and small ribosomal RNAs is shown below the blots.


A
16°C
23°C
30°C
37°C
WT
nrd1-5
nab3-11
nrd1-5/nab3-11
sen1-E1597K
B
SNR47
SNR39B
1500
1000
500
200
rRNA
*
*
sen1
nrd1
nab3
nrd1/nab3
WT

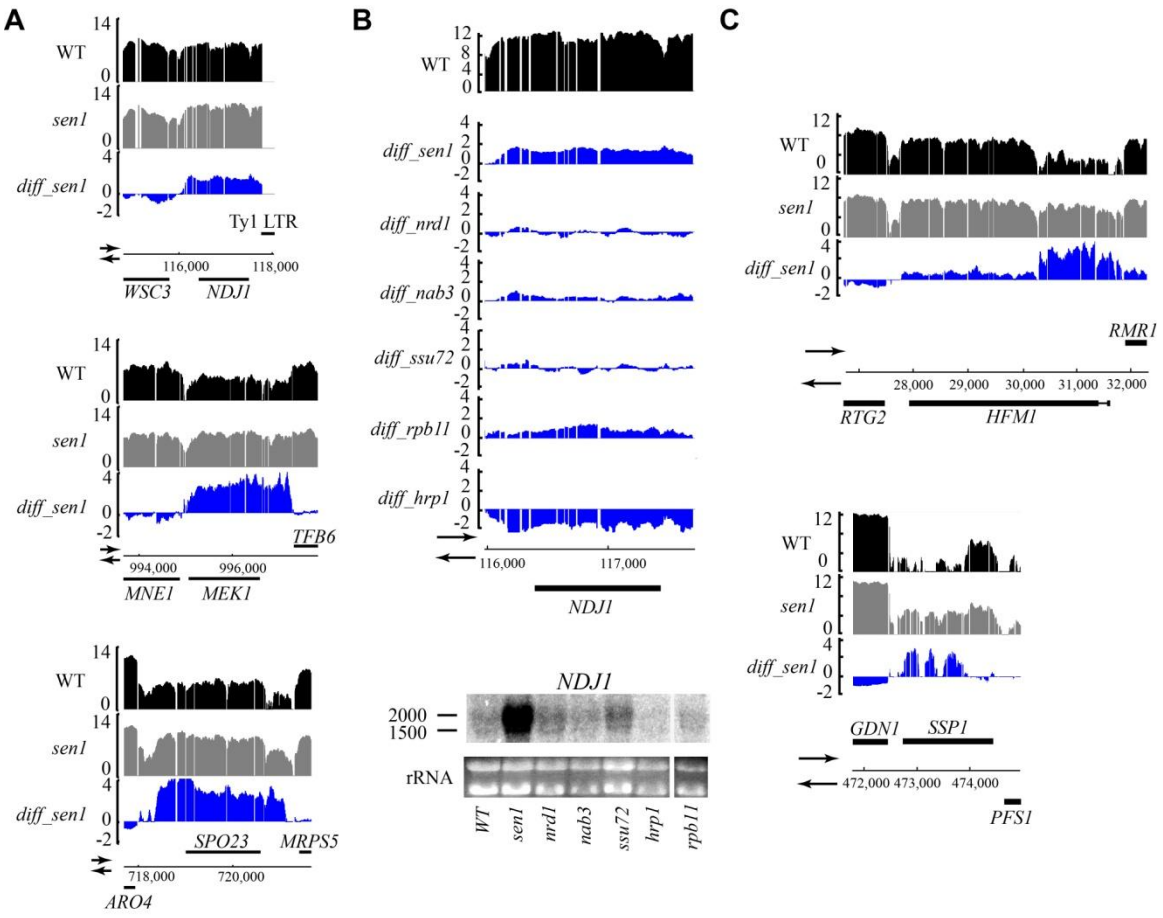
GO Term analysis (Holm-Bonferroni) for "cellular component" by Yeastmine of *Saccharomyces* Genome Database (SGD), the “meiotic cell cycle” GO term is the most enriched for up-regulated mRNAs with 79 matches (p value: 3.8e-14), and these up-regulated meiotic genes include *NDJ1, MEK1, and SPO1* (Table 4.4 and Figure 4.8A). In agreement with this finding, Sugiyama *et al.* reported that deletion of *SEN1* caused elevated mRNA level of some meiotic genes in fission yeast (SUGIYAMA *et al.* 2012). To confirm that the 79 identified meiotic genes are up-regulated during meiosis, I referred to the strand-specific gene expression profiling available by Saccharomyces Genomics Viewer (SGV) online (http://sgv.genouest.org/cgi-bin/viewer.cgi) (LARDENOIS *et al.* 2011). At least 71 of the 79 genes are increased during sporulation (Table 4.4).

Next, I performed stand-specific Northern blot analysis to validate the up-regulated genes identified from the transcriptome data. For example, the meiosis-specific protein Ndj1, encoded by *NDJ1*, accumulates at telomeres during meiotic prophase and promotes meiotic recombination (CONRAD *et al.* 1997; WU and BURGESS 2006). The *NDJ1* transcript is strongly increased during sporulation according to SGV and in the Sen1 mutant based on our transcriptome data (Figure 4.8B, top). Using a strand-specific oligo probe, I confirmed that the Sen1 mutation causes significantly increased accumulation of the *NDJ1* transcript (Figure 4.8B, bottom). Interestingly, Brar and colleagues have identified a regulatory upstream ORF (uORF) associated with *NDJ1* (BRAR *et al.* 2012), which could decrease the translation level of its downstream *NDJ1* ORF when present in the same transcript. However, due to the limited resolution, my Northern blot cannot distinguish the *NDJ1* mRNA from the potential uORF-*NDJ1* transcript. It is unclear whether mutations in *SEN1* induce accumulation of the 5'-extended transcript. Among the 79 mRNA genes from the GO term analysis, only *NDJ1*, *SPO23*, *IME2* and *TEP1* are uORF-associated (BRAR *et al.* 2012), and the low co-occurrence indicates the

TABLE 4.4 The 79 GO term "meiotic cell cycle" enriched mRNA genes that are over two fold up-regulated in the Sen1 mutant

Systematic Name	Standard Name	Increased in SGV	Systematic Name	Standard Name	Increased in SGV
YAL068C	*PAU8*	Y	*YHR184W*	*SSP1*	Y
YBR045C	*GIP1*	Y	*YIL017C*	*VID28*	N
YBR148W	*YSW1*	Y	*YIL073C*	*SPO22*	Y
YBR186W	*PCH2*	Y	*YIL120W*	*QDR1*	N
YBR250W	*SPO23*	Y	*YJL037W*	*IRC18*	Y
YCL048W	*SPS22*	Y	*YJL038C*	*LOH1*	Y
YCR010C	*ADY2*	Y	*YJL106W*	*IME2*	Y
YDL154W	*MSH5*	Y	*YLR035C*	*MLH2*	N
YDR076W	*RAD55*	N	*YLR054C*	*OSW2*	Y
YDR113C	*PDS1*	Y	*YLR213C*	*CRR1*	Y
YDR126W	*SWF1*	N	*YLR227C*	*ADY4*	Y
YDR218C	*SPR28*	Y	*YLR308W*	*CDA2*	Y
YDR260C	*SWM1*	Y	*YLR329W*	*REC102*	Y
YDR273W	*DON1*	Y	*YLR341W*	*SPO77*	Y
YDR285W	*ZIP1*	Y	*YLR343W*	*GAS2*	Y
YDR383C	*NKP1*	N	*YLR445W*	*GMC2*	Y
YDR402C	*DIT2*	Y	*YML128C*	*MSC1*	Y
YDR403W	*DIT1*	Y	*YMR017W*	*SPO20*	Y
YDR506C	*GMC1*	Y	*YMR133W*	*REC114*	Y
YDR522C	*SPS2*	Y	*YMR306W*	*FKS3*	Y
YDR523C	*SPS1*	Y	*YNL012W*	*SPO1*	Y
YEL072W	*RMD6*	N	*YNL128W*	*TEP1*	Y
YPL200W	*CSM4*	Y	*YNL196C*	*SLZ1*	Y
YER046W	*SPO73*	Y	*YNL210W*	*MER1*	Y
YER179W	*DMC1*	Y	*YOL047C*	*LDS2*	Y
YFL003C	*MSH4*	Y	*YOL091W*	*SPO21*	Y
YFL033C	*RIM15*	N	*YOL104C*	*NDJ1*	Y
YGL033W	*HOP2*	Y	*YOL132W*	*GAS4*	Y
YGL045W	*RIM8*	Y	*YOR178C*	*GAC1*	Y
YGL158W	*RCK1*	Y	*YOR190W*	*SPR1*	Y
YGL183C	*MND1*	Y	*YOR298W*	*MUM3*	Y
YGL249W	*ZIP2*	Y	*YOR313C*	*SPS4*	Y
YGL251C	*HFM1*	Y	*YOR351C*	*MEK1*	Y
YGR059W	*SPR3*	Y	*YPL027W*	*SMA1*	Y
YGR225W	*AMA1*	Y	*YPL121C*	*MEI5*	Y
YHL022C	*SPO11*	Y	*YPL130W*	*SPO19*	Y
YHR014W	*SPO13*	Y	*YPL164C*	*MLH3*	Y
YHR139C	*SPS100*	Y	*YPR007C*	*REC8*	Y
YHR079C-A	*SAE3*	Y	*YER044C-A*	*MEI4*	Y
YHR157W	*REC104*	Y	*YPL200W*	*CSM4*	Y

FIGURE 4.8 The E1597K substitution in Sen1 derepresses transcription of meiotic genes. A) Meiotic gene transcripts are accumulated in the Sen1 mutant. The transcriptome profiles around the meiotic genes *NDJ1*, *MEK1* and *SPO23* are shown here. The black traces show the absolute transcript levels in wild-type, the gray traces show the absolute transcript levels in the Sen1 mutant, and the blue traces show the differential transcript levels in the Sen1 mutant (all in log2 scale). The locations of the genes are indicated by the black bars. **B**) The Sen1 substitution causes increased accumulation of the *NDJ1* transcript. The top panel shows the transcriptome profiles of the region around *NDJ1*. The traces represent the same meanings as in A). The *NDJ1* gene is in the Crick strand indicated by the black bar. Northern blot analysis of *NDJ1* in wild-type and the six mutant strains is shown in the bottom panel. Total RNA from the indicated strains was resolved on denaturing agarose gels. The numbers to the left of the blots represent the length in nucleotides of RNA markers. Ethidium bromide stain of the filter containing the large and small ribosomal RNAs is shown below the blots. **C**) The 5' parts of the *HFM1* and *SSP1* transcripts that are repressed by Sen1 are accumulated in the Sen1 mutants. The transcriptome profiles around the regions of *HFM1* and *SSP1* are shown here. The traces represent the same meanings as in A).


A
WT
sen1
diff_sen1
Ty1 LTR
116,000
118,000
WSC3
NDJ1
TFB6
994,000
996,000
MNE1
MEK1
SPO23
MRPS5
718,000
720,000
ARO4
B
diff_nrd1
diff_nab3
diff_ssu72
diff_rpb11
diff_hrp1
116,000
117,000
NDJ1
2000
1500
rRNA
WT
sen1
nrd1
nab3
ssu72
hrp1
rpb11
C
RMR1
28,000
29,000
30,000
31,000
32,000
RTG2
HFM1
GDN1
SSP1
472,000
473,000
474,000
PFS1

process of Sen1-dependent meiotic gene repression is unlikely to be related to the uORF-based gene expression regulation.

Some genes exhibit a more significant increase in transcript from their upstream half than their downstream half. In wild-type, these genes are transcribed from an internal transcription initiation site, but start from an upstream initiation site in the Sen1 mutant, such as *HFM1* and *SSP1* (Figure 4.8C). Published strand-specific isoform transcript analysis confirmed that the 5' parts of their transcripts are absent in the vegetative cells (PELECHANO *et al.* 2014), and increased transcript levels of their entire coding regions are observed during sporulation (LARDENOIS *et al.* 2011). Therefore, decreased Sen1activity results in more uniform transcript levels across these genes.

4.9 Discussion

In this chapter, I investigate the functional relationships of different proteins involved in the Sen1-dependent transcription termination pathway. My discoveries from the whole transcriptome analysis reveal that the factors cooperate in the process of Sen1-dependent termination, but their functions are diverse, indicating that Pol II-mediated transcription termination is under complex, combinatorial control.

The transcriptome analysis using hypomorphic mutations overcomes the limitations of other genomic methods

Recently, there have been several studies on the Sen1-dependent termination pathway by different genome-wide approaches. In 2006, our lab performed the first genome-wide study of the Sen1 pathway applying ChIP-chip experiments on the *sen1-E1597K* mutant and wild-type strains and identified novel targets of this pathway (STEINMETZ *et al.* 2006b), but we only investigated a Sen1 substitution in this study. Later, several groups have applied high-throughput

RNA sequencing methods in combination with UV-crosslinking/immunoprecipitation or nuclear depletion of termination factors, to identify new targets and functions of Sen1-dependent termination (CREAMER *et al.* 2011; WLOTZKA *et al.* 2011; SCHULZ *et al.* 2013; SCHAUGHENCY *et al.* 2014). These studies have provided additional insights into the Sen1 pathway, but are restricted by their methods. By using the hypomorphic substitutions in the proteins involved in Sen1-dependent termination, we performed factor-specific, functional transcriptome analysis and revealed novel features of this pathway that were not uncovered by the other methods .

By the pull-down experiments coupled with RNA-seq, the distribution of the termination factors on RNA can be profiled, but this method alone cannot determine the function of the binding events, therefore tends to dismiss the underlying mechanistic differences. For example, similar enrichment patterns of Nrd1 and Nab3 were observed on both the *SNR47* and *SNR39B* transcripts (Figure 4.9A) (CREAMER *et al.* 2011), suggesting RNA binding abilities of Nrd1and Nab3 are needed for transcription termination of both snoRNA genes in the same way. However, our transcriptome data on *SNR47* and *SNR39B*, in concert with my later Northern blot analyses, argue against this idea, as these two snoRNA genes are apparently terminated by different mechanisms from the perspective of Nrd1 and Nab3 (see Section 4.7), in which the Nrd1 and Nab3 RRMs are both required for termination of *SNR47*, but neither for *SNR39B*.

Our transcriptome analyses using hypomorphic mutations suggested that, although substitutions of the six factors were all found to cause Sen1-dependent termination defects, different subsets of factors are required for transcription termination of different genes. The termination factor-nuclear depletion strategy by the anchor-away method can hardly be used to characterize the functional differences between termination factors due to lack of specificity. In the reported nuclear depletion experiment, Nrd1 or Sen1 were tagged with the FKBP12-

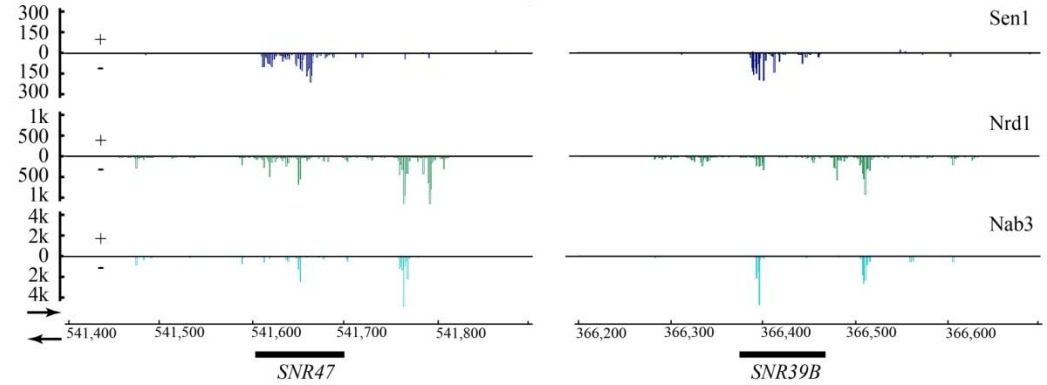
rapamycin-binding domain (FRB) in the yeast strains where the FKBP12 domain is fused to the ribosomal protein *RPL13A*. In the presence of rapamycin, Nrd1- or Sen1-FRB, *RPL13A*-FKBP12 and rapamycin form a ternary complex and are exported from the nucleus (HARUKI *et al.* 2008; SCHULZ *et al.* 2013; SCHAUGHENCY *et al.* 2014). However, as Nrd1, Nab3 and Sen1 form a ternary termination complex, depletion of one from the nucleus may decrease the nuclear abundance of the other two components, therefore causing non-specific effects.

To test the hypothesis about the specificity of the anchor-away method, I analyzed the results from the Nrd1 nuclear depletion experiment performed by (SCHULZ *et al.* 2013). In this study, they identified 32 mRNA genes as attenuated genes since their transcript levels are at least 1.25 fold up-regulated upon depletion of Nrd1. When I revisited them using our transcriptome data, I found only 13 of the 32 genes are up-regulated in the Nrd1 mutant. In contrast, 15 genes are up-regulated in the Sen1 or Nab3 mutant, but not the Nrd1 mutant. There are another four genes that are up-regulated in none of the Sen1, Nrd1 or Nab3 mutant (Figure 4.9B). Therefore, the 32 attenuated genes originally identified upon nuclear depletion of Nrd1are not all Nrd1-dependent. As the termination factors in a complex could be co-exported from the nucleus upon Nrd1 depletion, the up-regulated expression of the attenuated genes is likely to be caused by complex, combinatorial effects of disrupting the functions of Sen1, Nrd1, Nab3 and/or other unknown factors.

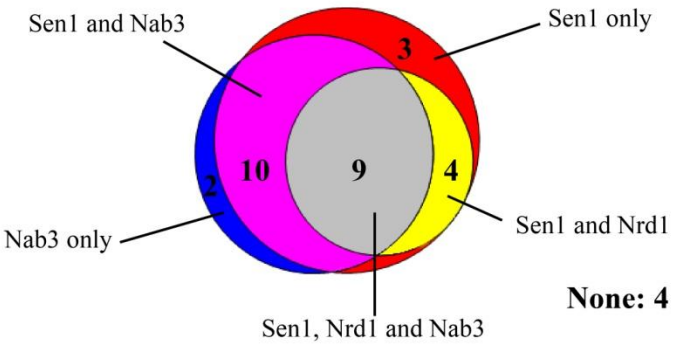
The specificity issue of the anchor-away method is also revealed by the sn/snoRNA termination analysis. Schulz *et al.* reported that 80% of sn/snoRNA genes showed termination defects upon nuclear depletion of Nrd1 (SCHULZ *et al.* 2013). I analyzed 48 Pol II-transcribed sn/snoRNA genes, only 25 of them (52%) showed termination defects in the Nrd1 mutant, while 42 genes (88%) exhibited termination defects caused by at least one of the six mutations.

FIGURE 4.9 Validation of other genomic studies of the Sen1-dependent termination pathway using our factor-specific functional transcriptome data. A) Occupancy of Sen1, Nrd1 and Nab3 on *SNR47* and *SNR39B* identified by PAR-CLIP (CREAMER *et al.* 2011). cDNA reads from the Watson and Crick strands are shown above and below the zero line respectively. *SNR47* and *SNR39B* are indicated by the black bars. Transcription units on the Watson stand are above the axis, and the Crick strand below the axis. **B)** Validation of the 32 reported attenuated genes based on the transcriptome data described in my thesis (SCHULZ *et al.* 2013). Genes that show at least 20% increase in transcript level in the indicated mutant according to my transcriptome data are considered as bona fide attenuated genes. Groups of genes that are identified as attenuated in indicated mutants are shown using the Venn diagram and separated by colors. There are four genes that are identified as attenuated genes in none of the three mutants. The Venn diagram is generated by Venn Diagram Plotter (Pacific Northwest National Laboratory, http:// http://omics.pnl.gov/software/venn-diagram-plotter).

A.


300
150
0
150
300
Sen1
1k
500
0
500
1k
Nrd1
4k
2k
0
2k
4k
Nab3
541,400
541,500
541,600
541,700
541,800
SNR47
366,200
366,300
366,400
366,500
366,600
SNR39B

B.


Sen1 and Nab3
Sen1 only
3
10
9
4
2
Nab3 only
Sen1 and Nrd1
None: 4
Sen1, Nrd1 and Nab3

Therefore, the termination defects of sn/snoRNA genes upon Nrd1 depletion are also partly caused by the non-specific effect of the anchor-away method.

In addition, the depletion efficiency may vary on different target proteins. For example, although Sen1 has an essential function in the nucleus, nuclear depletion of Sen1 did not cause lethality (SCHAUGHENCY *et al.* 2014).

By using known hypomorphic mutations for the transcriptome analysis, factor-specific effects across the transcriptome can be identified, to further elucidate the functions and relationships of different termination factors.

Differential factor requirement for Sen1-dependent termination of different genes

Although the six proteins characterized in this chapter are all involved in Sen1-dependent termination, the distinct correlation coefficients (0.08-0.66) revealed by the genome-wide correlation analyses indicate the target genes that are affected by the mutants only partially overlap, therefore the involvement of different factors in this termination pathway varies. Further, by sn/snoRNA read-through ratio analysis of the six mutants, I confirmed the differential factor requirement for termination of different genes, which is likely to represent the situations of other groups of genes terminated by the Sen1 pathway.

Among the six Sen1 termination components, Sen1 is required for efficient termination of more than 80% tested sn/snoRNA genes, followed by Nrd1, Nab3 and Ssu72, which function on at least two thirds as many sn/snoRNA genes as Sen1. Rpb11 and Hrp1 are also required for sn/snoRNA gene termination but much less often, as their substitutions affect less than 30% of the analyzed genes. Although these proteins are differentially required for termination of individual genes, Sen1 appears to play an indispensible role in this termination pathway, as the Sen1 activity is required for efficiently terminating the majority of short sn/snoRNA genes,

including all the identified targets of other factors involved in this pathway.

My studies add a novel perspective of the functional coordination of Nrd1 and Nab3 on the prevailing model of the Sen1 termination complex, where Nrd1 and Nab3 are proposed to function in the form of heterodimer to recruit Sen1 for transcription termination. My results showed that the substitution in the RRM of Nrd1 or Nab3 only affects termination of a subset of Sen1's targets, and the targets of Nrd1 and Nab3 are not identical (see Section 4.6). Moreover, in the analyses of *SNR47* and *SNR39B* using the double mutant XCY377, I found Nrd1 and Nab3 cannot fully compensate for each other's function on *SNR47*, and disrupting the RNA-binding abilities of both does not affect Sen1-dependent termination of *SNR39B*. Although genome-wide analyses are necessary to clarify the relationship between the distribution of mutant Nrd1 and Nab3 and changes of transcription termination caused by their mutations, my results suggest an alternative way for Sen1 recruitment which does not require the heterodimer of Nrd1 and Nab3.

Sen1-dependent termination represses meiotic gene expression

Meiosis is a highly regulated process in response to dynamic changes in gene expression, which serve as one molecular basis of the programmed remodeling events in meiotic cells. In Section 4.8, I showed Sen1-dependent termination may regulate expression of meiotic genes by transcription repression in vegetative cells, either the entire gene or only the 5' part of the gene. This function of Sen1 may be important for cells to enter meiosis, or even specific meiotic phases. For example, the repressed upstream half of *HFM1* by Sen1 encodes its DExH box DNA helicase domain, and the DNA helicase activity of Hfm1 functions for yeast meiotic crossing over (NAKAGAWA and OGAWA 1999; NAKAGAWA and KOLODNER 2002). In vegetative cells, the internal initiation allows expression of the truncated downstream half of *HFM1* which may be indispensable, while Sen1-dependent termination represses the expression of the upstream

meiotic helicase region. Perhaps inactivating Sen1 at the termination site of the *HFM1* upstream half would turn on the helicase activity for meiotic crossing over. However, whether the uniform transcript level across the gene in the Sen1 mutant represent the contiguous transcript need to be confirmed first.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

In the preceding chapters of this dissertation, I have presented the results of my experiments that expand the understanding of the mechanism and targets of Sen1-dependent transcription termination.

In Chapter 3, my genetic analyses provided new evidences for the essential role of the Sen1 helicase domain in cell viability and transcription termination. I found that the helicase domain plus a nuclear localization sequence of Sen1 is the minimal region sufficient for yeast cell viability. The Sen1 helicase domain is essential for efficient transcription termination, and the E1597K substitution that disrupts an intradomain ionic bond of the helicase domain results in strong termination defects. By assessing 13 missense mutations in the Sen1 helicase domain that co-segregate with the neurodegenrative disease AOA2, I have shown that five of the mutations are recessive lethal and cause dominant terminator read-through effects. Furthermore, termination defects appear to be a common feature of these AOA2-associated mutations, as 10 of them cause read-through of at least one of the three tested terminators.

In Chapter 4, I identified the diverse effects across the transcriptome caused by hypomorphic mutations in Sen1, Nrd1, Nab3, Ssu72, Rpb11 and Hrp1, which are all factors involved in Sen1-dependent termination. By investigating the effects of the six mutations on the entire transcriptome as well as termination of sn/snoRNA genes, I found the Sen1 substitution induces the broadest and strongest termination defects, while the other substitutions only affect termination of different subsets of Sen1's targets. Thus, Sen1-dependent termination of different genes exhibits variable requirements for termination factors. A following test using a Nrd1-Nab3 double mutant revealed that Nrd1 and Nab3 are not functionally redundant. I also identified a potential regulatory role of Sen1-dependent termination in repressing meiotic genes in vegetative cells.

With these discoveries, there are new questions to be addressed, such as: What is the mechanism that the Sen1 helicase domain uses to terminate transcription? Which termination factors are required for termination of different Sen1 target genes? Below I discuss potential experiments to answer those questions.

5.1 The molecular basis of the Sen1 helicase domain for transcription termination

The crystal structure of the Sen1 helicase domain would elucidate important features of it at the molecular level, and facilitate our understanding of its structure-function relationships. In Chapter 3, I showed that the crystal structure of the yeast Upf1 helicase domain can be used as a reliable model to predict the functions of the conserved residues in Sen1. However, the large insertion domains, 1B and 1C, are dissimilar between Upf1 and Sen1, and functionally essential, as deleting or replacing them with Senataxin 1B and 1C domains cannot confer viability (Section 3.5). Therefore, the crystal structure of the Sen1 helicase domain is necessary. Toward this goal, our lab has successfully purified the recombinant Sen1 helicase domain (Sen1-HD) (residues 1095-1876) from *E. coli* and improved its yield by a codon-optimized construct (S. Martin-Tumasz and E. Montemayor, unpublished data). This recombinant protein could be tested to crystallize in a complex with nucleic acids and a nucleotide analog, and the results can provide structural insights into the functions of the Sen1 helicase domain.

The recombinant Sen1-HD allows *in vitro* biochemical studies of the Sen1 helicase domain. Our lab has proved that it has ATP-independent single-stranded (ss) DNA and RNA binding ability, and translocase activity on both in a 5' to 3' direction in the presence of ATP (S. Martin-Tumasz and D. Brow, unpublished data). I have presented that mutations in the Sen1 helicase domain can affect Sen1-dependent termination *in vivo*, and proposed that the

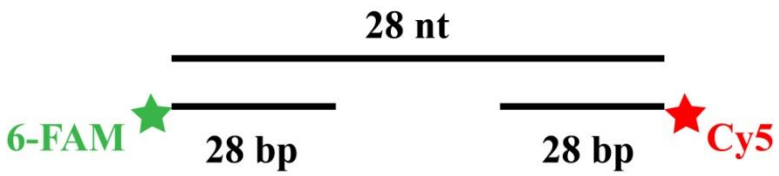
termination defects are caused by disrupting the helicase-related activities based on the Upf1 structure model (Sections 3.6 and 3.8).

To further determine how the helicase domain functions in transcription termination, the Sen1 helicase domain variants containing the point mutations that I investigated *in vivo* could be characterized *in vitro* using the assays identical with that for wild-type. For example, the T1779P mutation is expected to interfere with the nucleic acid binding ability, and it could be tested for the nucleic acid binding and translocase activities. The binding ability could be tested on single-stranded DNA and RNA in the absence of ATP. For the helicase assay, our lab synthesized the nucleic acid trimeric complexes (Figure 5.1) that have 6-FAM- and Cy5-labeled, 28 nucleotides long DNA or RNA oligos annealed to the 5' and 3' ends of an 84 nucleotide long DNA or RNA oligo. In the presence of ATP, the wild-type Sen1-HD and the variant could be incubated with different nucleic acid trimer combinations. The helicase directionality and activity could be determined based on the reaction products. Another mutant, *sen1-R1820Q*, is proposed to disrupt the hydrogen bond between R1820 and the ATP γ-phosphate, and affect the ATPase activity. The nucleic acid-dependent ATPase activity of this variant could be tested by measuring its ability of converting [α-32P]ATP to [α-32P]ADP in the presence or absence of short RNA or DNA olionucleotides (PORRUA and LIBRI 2013). The *in vitro* biochemical assays, along with the results from the *in vivo* genetic studies, would verify the proposed molecular defects caused by the mutations and the roles of Sen1 helicase activities in transcription termination.

5.2 Developing the *in vitro* reconstituted transcription termination system

The mechanism underlying Sen1-dependent termination could be studied by the *in vitro* reconstituted transcription termination system. Porrua and Libri have initiated reconstitution of a simplified Sen1 transcription termination system *in vitro*. Using the full-length Sen1 purified

FIGURE 5.1 Schematic diagram of the nucleic acid trimeric complex. The black lines represent single-stranded nucleic acids. Their lengths are indicated. The short oligo annealed to the 5' end of the long oligo is labeled with a 6-FAM (green) fluorophore at its 3' end. The short oligo annealed to the 3' end of the long oligo is labeled with a Cy5 (red) fluorophore at its 3' end. Complexes of all possible combinations of DNA and RNA were generated.


28 nt
6-FAM
28 bp
28 bp
Cy5

from yeast, they showed that Sen1 could dissociate RNA and Pol II molecules from the DNA template in an *in vitro* transcription reaction in an ATP-dependent manner upon artificially stalling Pol II by the Reb1protein (PORRUA and LIBRI 2013).

This simplified system needs to be optimized to more closely represent the *in vivo* situation, and meet the needs of testing more conditions and Sen1 termination factors in it. For example, a major limit of this system is the artificial change of Pol II's elongating rate by using the road-block Reb1. The prevailing model suggests that the phosphorylation state transition of the CTD from the Ser5-P enrichment to the Ser2-P enrichment mediates the kinetic change of Pol II relative to Sen1, and allows termination to occur (see Section 1.2). Therefore, to improve the *in vitro* system, the next step would be to verify the CTD-mediated Pol II kinetic change and establish it in the *in vitro* system. The resulting system would provide a powerful way to determine the functions of the CTD phosphorylation states, CTD-binding termination factors, Sen1 helicase domain mutations and *cis*-acting elements on RNA in the process of Sen1-dependent termination.

5.3 Mutation-specific effects revealed by yeast Sen1 suggest a strategy for determining the cause of AOA2

Previously, Fogel *et al.* showed two *SETX* mutations, one associated with AOA2 and the other one associated with ALS4, caused changes in expression of distinct sets of genes when the mutant alleles were overexpressed in senataxin-haploinsufficient fibroblasts (FOGEL *et al.* 2014). This finding indicates mutation-specific defects of *SETX* on gene expression underlying the difference between AOA2 and ALS4. In Chapter 3, I described my characterization of 13 AOA2 mutations in yeast Sen1. These mutants exhibited different effects on cell growth and termination, thus, mutation-specific effects could exist between different AOA2 patients.

Several groups have used cells derived from AOA2 patients to identify the defects caused by the disease mutations in human. Using fibroblasts and a lymphoblastoid cell line derived from two patients carrying different AOA2 mutations, Suraweera *et al.* showed that AOA2 cells exhibited increased sensitivity to DNA damage reagents and defective double-strand break (DSB) repair (SURAWEERA *et al.* 2007). Using the same AOA2 patient cell lines, Miller *et al.* identified higher expression of several genes encoding antiviral products from the AOA2 cells than that of cells from healthy donors upon virus infection (MILLER *et al.* 2015). These studies identified defects that are associated with AOA2 cells. However, partially due to the small number of patient samples analyzed, they could hardly reveal which genes and pathways, when their expression levels are altered by the disease mutations, directly contribute to the pathogenesis of AOA2. The AOA2-associated mutations are diverse, including missense mutations, nonsense mutations, frameshift mutations, indels, and splicing site mutations. The diversity of disease mutations, together with the mutation-specific effects as suggested by my studies in yeast Sen1 and the varied physiological conditions between patients, could alter gene expression to highly distinct patterns between patient samples. Thus, it is important to identify the common defects underlying AOA2 for determining the cause of this disease. AOA2 is a rare disease, only less than 200 patients have been reported, and the number of genotypes is even less. This fact enables genome-wide gene expression profiling of patient samples of all available genotypes. By comparing the expression profiles, genes and pathways that are altered most frequently throughout all the samples are likely to contribute to the pathogenesis of AOA2.

5.4 Optimizing the method for the transcriptome analysis

In Chapter 4, I analyzed the functional transcriptome data from the microarray experiment, and revealed diverse effects caused by hypomorphic mutations in six factors

involved in Sen1-dependent termination. Because the microarray data alone cannot discriminate transcripts from different strands, a significant amount of result validation work is required. To expedite the research, I propose performing strand-specific RNA-seq experiments on the current six mutants, their isogenic wild-type strains and the new Nrd1-Nab3 double mutant XCY377. The temperature shift treatment would be applied to all the samples. The strand- and factor-specific transcripts and changes of transcript levels across the genome would be profile by this approach.

Although the regular RNA-seq approach would be sufficient to identify changes of the transcript level, certain optimization of the RNA-seq method could be made specifically for analyzing transcription termination. The transcript isoform sequencing (TIF-seq) approach identifies the exact 5'- and 3'- boundary of each transcript (PELECHANO *et al.* 2014). Therefore, this method has the advantage of distinguishing transcripts associated with termination defects, such as polycistronic transcripts and overlapping transcripts that differ in length. Meanwhile, to improve identification of the factor-specific effects that are enhanced by a temperature shift, 4-thiouracil (4tU) could be used to label newly-synthesized transcripts after the temperature shift, and the labeled transcripts could be purified and sequenced to identify effects specifically caused by the mutations upon the temperature shift (SCHULZ *et al.* 2013). In addition, 4tU labeling does not affect normal yeast cell physiology, and allows growth of yeast in the YEPD medium (SUN *et al.* 2012).

5.5 Characterizing non-canonical targets of Sen1-dependent termination

In the prevailing model of Sen1-dependent termination, the heterodimer of Nrd1 and Nab3 can recognize and bind to their consensus sequences on the nascent transcript, and recruit Sen1 to the transcript to terminate transcription. In Chapter 4, I presented my results that expand

this, as only a subset of Sen1's targets requires the RNA-binding ability of Nrd1 and Nab3. The differential requirement for termination factors of different genes suggests Sen1-dependent termination is under a complex control. Therefore, I propose the experiments to characterize the non-canonical targets of Sen1-dependent termination and identify the mechanism underlying the complex control. Transcription termination on *SNR39B* could be an ideal model gene for this analysis, since disrupting the RNA-binding ability of Nrd1 and Nab3 simultaneously did not affect efficient termination of this gene. It should be noted that there could be several classes of Sen1 target genes utilizing different mechanisms. Nevertheless, the studies using *SNR39* as a model would establish a framework for characterizing other non-canonical targets of Sen1-dependent termination.

Based on the position of its downstream poly(A) sites, the primary transcript of *SNR39B* appears to be no more than 320 nucleotides long . There are several consensus sites of Nrd1 and Nab3 within this region (Figure 5.2), which may contribute to the interactions between *SNR39B* and the two RNA-binding proteins identified by the RNA cross-linking experiment (Figure 4.9). To map the *cis*-acting elements on *SNR39B* that are important for termination, the downstream 222 bp sequence could be inserted into the *CUP1* reporter construct for the copper resistance assay. Random mutagenesis by error-prone PCR on this region could be performed to identify which parts of this region are essential for terminating *SNR39B* (STEINMETZ *et al.* 2006a). Since there are consensus sequences for the Sen1 helicase domain, Nrd1, Nab3 and Hrp1 binding within this region (Figure A5.2), the identified key sites that are important for termination of *SNR39B* could be compared with the consensus sequences, to determine if any of the interactions conferred by the consensus sequences and the termination factors are involved in transcription termination of *SNR39B.*

FIGURE 5.2 Sequence of *SNR39B* and its downstream 222 bp region. The mature *SNR39B* sequence is highlighted in underlined bold font. The consensus sequences of Nrd1 and Nab3 are highlighted in green and yellow, respectively. The CAN repeats that may bind to the Sen1 helicase domain are highlighted in red. The regions that encode the UA-repeat sequences for Hrp1 docking are highlighted in cyan.

ATTTAGTGATGATACTGCCGATTCTGGCATTCCAAAAAGTGACTAGCA

AAATTGCGATGTTGTCAACTTAAATTACACCATCTTTCGGGGCTGATA

TTTAATACATCTATTTTATCCATTCTTCTACGTATTACGTTTGTAGCTTGGC

CAAAAAAATATCTTGCATAATTTGTACTCCAGTTCACATCAGTTGAGTAGA

GAATAGCTACTGATCATATAAACACCCACACATATCTATAAATCCGTACC

TTTCCTATATATATATATATATATTTTTTCAAATTTTGTCTGAAATTCTTTCTG

CAGGTTTGTCGCTGCCG

After the *SNR39B* terminator is determined, I could perform random mutagenesis by error-prone PCR on Nrd1, Nab3, Hrp1 and Sen1 to see if any mutations in them could cause read-through of the *SNR39B* terminator. Random mutagenesis could be performed across Nrd1, Nab3 and Hrp1, and the purpose is to determine whether they are involved in termination of *SNR39B*, and further the involvement is in an RNA-binding dependent or independent manner. Since the substitution E1597K in the helicase domain caused termination read-through of *SNR39B*, mutagenesis in Sen1 could be conducted in the regions outside its helicase domain. If read-through mutations in *SEN1* are identified in those regions, the location of those mutations would implicate the termination factor-interacting sites in Sen1 for *SNR39B*. The Sen1 mutants that exhibit temperature sensitive phenotypes are particularly valuable, because high copy number yeast genome library could be transformed into the mutants to identify genes that can suppress such growth defects. Those genes are likely to be involved in termination of *SNR39B*. Alternatively, the wild-type cells containing the *SNR39B-CUP1* reporter construct could be incubated on medium containing copper at permissive temperature (30 ℃) to select for spontaneous read-through mutants. If the selected mutants are temperature-sensitive, genes bearing the unknown read-through mutations could be identified by transforming the yeast genome library into the viable strain and look for complementing clones (STEINMETZ and BROW 2003; STEINMETZ *et al.* 2006a).

Screens for unknown termination factors could also be performed on a large-scale genetic platform. The barcoded yeast mutant libraries of deletion of non-essential genes and conditional temperature-sensitive alleles of essential genes have been constructed by other groups (PIERCE *et al.* 2007; LI *et al.* 2011). These libraries could be introduced into the isogenic strains of 46**a** or 46**α** that are designed for the *CUP1* reporter assay (see Section 2.2). The *SNR39B- CUP1*

reporter construct could be transformed into the new barcoded yeast strain library, and read-through mutants could be selected for by incubating cells on medium containing copper at the permissive or semi-permissive temperature. Ideally, the new barcoded yeast strain library would be a useful development of the *CUP1* reporter system and facilitate studies of other aspects of RNA biogenesis.

5.6 Investigating the function of Sen1-dependent termination in meiosis

In Chapter 4, I showed that Sen1 could repress meiotic gene expression in vegetative cells, implicating a regulatory role of Sen1 in yeast cell cycle. To understand how Sen1-dependent termination regulates yeast meiosis, the first step would be to systematically identify the meiotic transcripts that are targeted by Sen1. Brar *et al.* have mapped the yeast meiotic program in high resolution by both RNA-seq and ribosome profiling methods, which revealed strong, meiotic stage-specific expression of genes (BRAR *et al.* 2012). Their RNA-seq results could be analyzed in together with our transcriptome profiling of the Sen1 mutant, to determine which meiotic genes are targeted by Sen1, and further specify the stages when their transcripts are up-regulated. Meanwhile, unannotated transcripts in yeast appear to have a significant impact on new protein synthesis. Small open reading frames (ORF) have been shown to be widely present in unannotated transcripts in yeast (SMITH *et al.* 2014), and Brar *et al.* identified a strong, meiosis-specific shift of ribosome footprints to the unannotated ones , which may indicate the small ORFs are actively translated into new functional peptides during meiosis (BRAR *et al.* 2012). Therefore, in addition to the annotated meiotic genes, the potential function of Sen1 on biogenesis of the new meiotic peptide could also be investigated by profiling the ribosome footprints on meiotic transcripts that are repressed by Sen1-dependent termination.

Upon identification of the meiotic transcripts that are repressed by Sen1-dependent termination, their terminators could be characterized and tested in the *CUP1* reporter assay. The genetic method identical to that in Section 5.5 could be used to find termination factors that collaborate with Sen1 for meiotic gene repression. The meiotic gene-specific termination factors may be involved in meiosis-specific signaling pathways.

APPENDIX 1: GENETIC SELECTIONS FOR NEW TERMINATOR READ-THROUGH MUTATIONS IN THE SEN1 HELICASE DOMAIN

A1.1 Introduction

In Chapter 3, I used the *CUP1* reporter assay in the plasmid-shuffle yeast strain DAB206 to evaluate terminator read-through effects caused by *SEN1* mutants. In an attempt to investigate the regions of Sen1 that do not have sufficient structure information, I performed genetic selections for terminator read-through mutations, by combining error-prone PCR with the *CUP1* reporter assay. I conducted three random mutagenesis selections, using different terminator-*CUP1* reporter constructs and targeting different regions inside the Sen1 helicase domain. In total, 12 copper resistant strains probably bearing read-through mutations in *SEN1* were obtained from the selections.

A1.2 New terminator read-through mutations in the C-terminal half of the Sen1 helicase domain

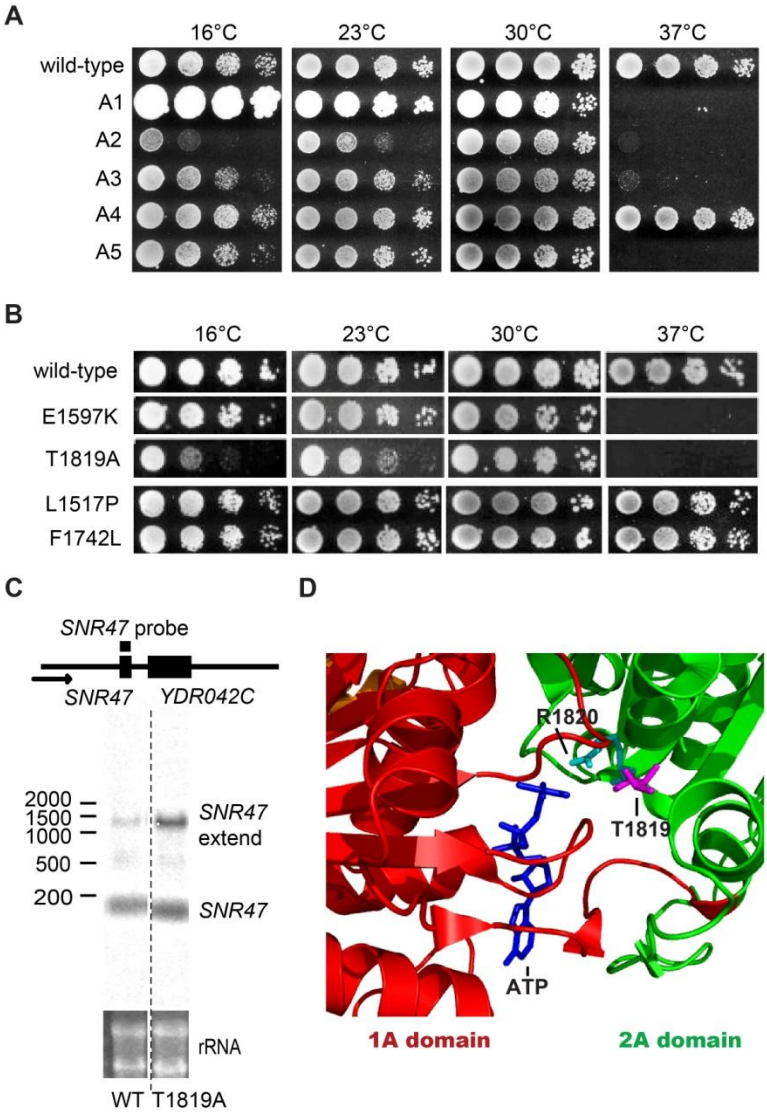
In the first selection, I looked for mutations in the C-terminal half of the helicase domain that cause the *CYC1* terminator read-through. Five copper resistant strains (numbered A1-A5 in the order of appearance on the copper plate) were obtained. Interestingly, the appearance of A1, such as its colony size and color, was distinct from the others.

To identify the potential read-through mutations, I PCR-amplified the *SEN1* fragment that includes the mutagenized region from genomic DNA of the selected strains and sequenced the PCR products. The PCR product sequences of all the strains could be aligned back to the correct positions in *SEN1* except the A1 strain. Although the PCR product from the genomic DNA of A1 had a similar size to others, its sequence did not match the *S. cerevisiae* genome. By sequence BLAST, I found its entire sequence unambiguously matches a region in the DEHA2G20328p mRNA (*SSD1*) gene of the yeast *Debaryomyces hansenii* CBS767 with 98% identity, while no significant similarity was found between the A1 sequence and the *S. cerevisiae*

SSD1. I compared the pair of oligos (SEN1-3F and SEN1-5R, see Table 2.1) that were used to amplify the yeast genomic DNA with the sequence of *D. hansenii SSD1* gene. In fact, each oligo has an over 10 nucleotide stretch that has a perfect sequence match in *D. hansenii SSD1*, and the distance between them is about 1.8 kb, close to the size of the correct PCR product from *SEN1* (~1.9 kb). I confirmed the growth of the A1 strain on medium containing copper. *D. hansenii* is extremophilic yeast, has been used for the synthesis of diary product and meat fermentation, and found in cheese, beer, meat, fruit, water, soil, air, *etc* (BREUER and HARMS 2006; GROENEWALD *et al.* 2008). The property that *D. hansenii* grows at low temperature but not at 37 ℃ (NGUYEN *et al.* 2009) is consistent with the growth phenotype of A1 (Figure A1.1A). I speculate A1 to be a copper-resistant *D. hansenii* strain that contaminated my copper plate, probably during replica plating.

The strains A2 to A5 showed varied growth phenotypes at different temperatures (Figure A1.1A). A2 showed severe cold- and heat-sensitivity. A3 and A5 are both lethal at 37 ℃, but only showed weak growth defects at 16 ℃. A4 exhibited no growth defect at all the tested temperatures. No missense or nonsense mutation was found in the mutagenized region of A3 and A4, while A2 and A5 each have two missense mutations there (A2: K1761R and T1819A, A5: L1517P and F1742L) (Figure 3.4). Although A5 exhibited strong heat-sensitivity, neither of its substitutions alone confers this defect (Figure A1.1B). I speculate that substitutions of these two hydrophobic residues may interfere with proper folding and stability of Sen1, and exhibit a synthetic effect. The strain A2 is both cold-and heat-sensitive, and I confirmed the substitution T1819A alone can render this growth phenotype (Figure A1.1B). Notably, cold-sensitivity has not been observed in any other missense mutation of *SEN1*. Further tests revealed this mutation causes strong read-through effects of all the tested terminators (Table 3.2), as well as termination

FIGURE A1.1 Characterization of the selected Sen1 T1819A mutation. A) Growth phenotypes at 16, 23, 30 and 37℃ of all the five mutants (A1 to A5) from Selection I. **B)** Growth phenotypes at 16, 23, 30 and 37℃ of the haploid wild-type and mutants harboring the E1597K or T1819A substitution after eight-fold serial dilution on YEPD medium. **C)** Total RNA from haploid strains harboring the wild-type or *sen1-T1819A* alleles was resolved on denaturing agarose gels and detected by the *SNR47* probe. The amplicon used as the probe is indicated by the bar above the gene map, and the arrow marks the transcription orientation. The numbers to the left of the blots represent the length in nucleotides of the RNA marker bands. The positions of the *SNR47* and *SNR47-YDR042C* read-through transcript (*SNR47*-extend) bands are indicated. Ethidium bromide stain of the portion of the filter containing the large and small ribosomal RNAs is shown below the blots. **D)** Locations of Sen1 T1819 and R1820 modeled on the Upf1 helicase domain. Sen1 T1819, Sen1 R1820 and ATP are shown in stick.


A
16°C
23°C
30°C
37°C
wild-type
A1
A2
A3
A4
A5
B
16°C
23°C
30°C
37°C
wild-type
E1597K
T1819A
L1517P
F1742L
C
SNR47 probe
SNR47
YDR042C
2000
1500
1000
500
200
SNR47 extend
SNR47
rRNA
WT T1819A
D
R1820
T1819
ATP
1A domain
2A domain

defects of *SNR47* at its genomic locus (Figure A1.1C). According to the Upf1 structure model (Figures 3.2 and A1.1D), this mutation is inside the ATP binding pocket and hydrogen-bonds to the ATP γ-phosphate-contacting residue R1820 of the Sen1 helicase domain. I hypothesize this T1819A substitution can change the ATP binding pocket and interfere with ATP binding/hydrolysis.

A1.3 New terminator read-through mutations in the 1B and 1C insertion domains

In the second and third selection, I used the *SNR47* terminator to select for read-through mutations in the insertion domains 1B or 1C. Seven strains were selected from 2,000 colonies (numbered B1-B3 and C1-C4 in the order of appearance on the copper plate), whereby B1-B3 are from the 1B domain mutagenesis and C1-C4 from the 1C domain mutagenesis.

To identify the potential terminator read-through mutations, I PCR-amplified fragments of *SEN1* that included the mutagenized regions from genomic DNA of the selected strains and sequenced the PCR products. C1 and C2 have the identical mutation, indicating they are siblings. Otherwise, the strains have distinct mutations. Surprisingly, each strain has a nonsense mutation, and all three types of stop codons are present (Table A1.1). Since we have identified that the intact helicase domain is essential for viability (Figure 3.2B), the nonsense mutant alleles of *SEN1* must be translated through these premature stop codons in order to maintain the required helicase domain.

I chose the W1166X mutation as a proof of concept to test the hypothesis, as it is the most upstream nonsense mutation, and will have the shortest translation product and the least viability if no translation read-through happens. This mutation coincides with the AOA2 disease substitution W1166S on the same residue, which was identified as a read-through mutation in Chapter 3 (see Table 3.2). DAB206 strain containing the *sen1-W1166X* allele created by site-

directed mutagenesis was confirmed to be viable, and exhibited growth defects at both 16 ℃ and 37 ℃ (Figure A1.2A). A band corresponding to the full-length Sen1 protein was observed in an immunoblotting assay (Figure A1.2B), which strongly supports the translational read-through hypothesis. This mutation caused read-through of both *SNR47* and *CYC1* terminators according to the copper resistance assay (Figure A1.2C), but I did not observe apparent termination defects at the genomic *SNR47* locus (Figure A1.2D). In addition, the *sen1-W1166X* mutation, as well as the *sen1-E1597K* mutation, could increase the expression level of *CUP1* that are fused to the downstream of the *SEN1* promoter/5'-UTR region (positions -434 to -1, relative to +1 transcription start site) (Figure A1.2E), suggesting transcription of *SEN1* could be auto-regulated by the *cis*-acting elements in its promoter/5'-UTR region. It would be interesting to test whether the *SEN1* attenuator construct can be read-through by other mutations in the Sen1-dependent termination factors.

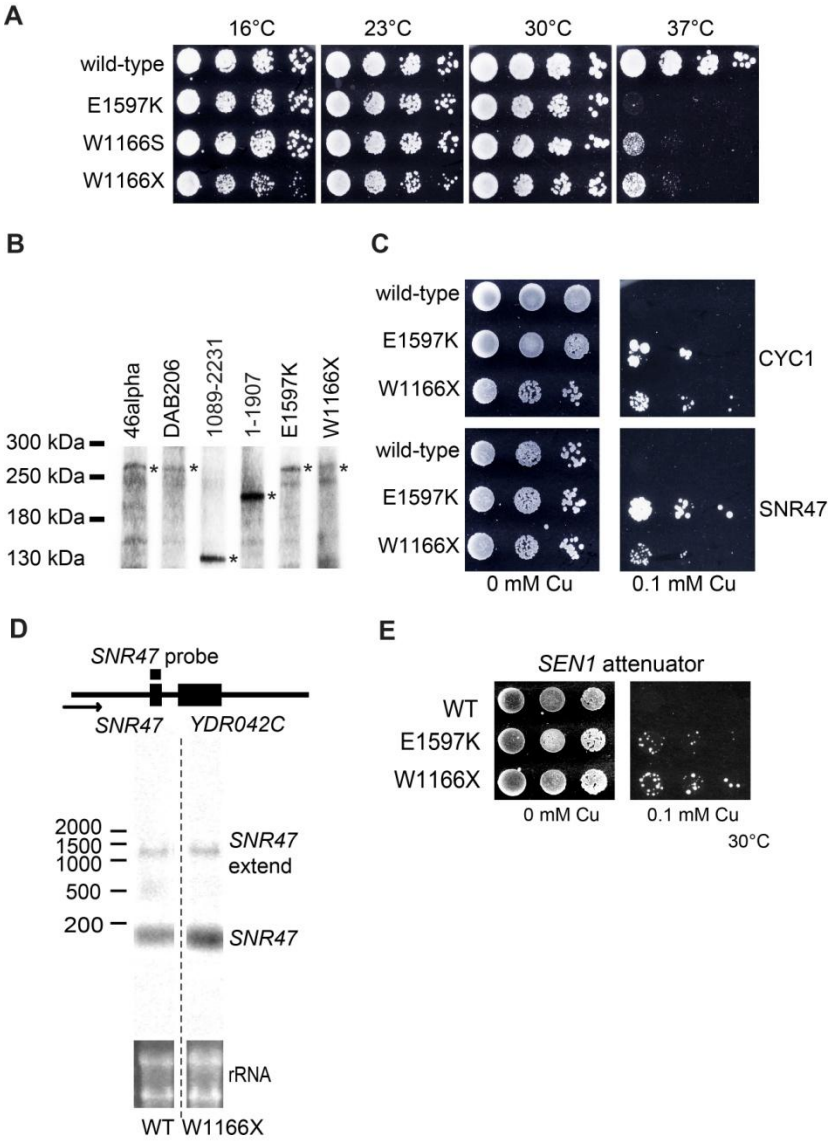
Three stop codon "recoding" mechanisms that allow in-frame translation read-through have been defined (GESTELAND and ATKINS 1996; STEINMETZ and BROW 1998): (1) insertion of a standard amino acid residue to the stop codon by near-cognate tRNA; (2) "translational bypassing" of stop codon due to large-scale slippage of the mRNA with respective to the peptidyl-tRNA specified by the preceding codon; and (3) insertion of selenocysteine to certain UGA codons. Because all three types of stop codons are found in the selected mutants, the first two mechanisms are more plausible to explain the viable phenotypes of the Sen1 nonsense mutants. The resulting missense or deletion mutations may render the translation read-through products decreased in function for transcription termination. Another possibility is that the truncated Sen1 peptides due to premature translation termination confer dominant negative effects on efficient transcription termination and inhibit the function of the full-length Sen1.

TABLE A1.1 Summary of the read-through mutants from Selection II

No.	Nucleotide change [a]	Codon change	Other codon changes
B1	G3497A (TAG)	W1166X	K1284R
B2	A3751T (TAA)	K1251X	R1230G, K1238K, R1263R
B3	T3585G (TAG)	Y1195X	C1209Y
C1	A4468T (TGA)	R1490X	V1486G, R1531R
C2	A4468T (TGA)	R1490X	V1486G, R1531R
C3	A4651T (TGA)	R1551X	A1451A, I1465T, N1527S, R1531R
C4	C4700A (TAG)	S1567X	R1531R, I1564I, L1569L

[a] The nucleotide change that leads to a premature stop codon and the resulting stop codon.

FIGURE A1.2 Characterization of the selected Sen1 nonsense mutation W1166X. A) Growth phenotypes at 16, 23, 30 and 37 ℃ of the haploid wild-type and mutants harboring E1597K, W1166S or W1166X substitution after eight-fold serial dilution on YEPD medium. **B)** Anti-Sen1(HD) immunoblot of cell extracts from haploid strains that contain either the wild-type allele (46alpha and DAB206) or the indicated *SEN1* alleles. Bands with the expected apparent molecular mass are indicated with asterisks. **C)** Copper resistance assay for terminator function. Strains containing the indicated *SEN1* alleles and either the CYC1 or SNR47 reporter construct were serially diluted eight-fold, spotted on synthetic complete medium with no or 0.1 mM copper sulfate and incubated at 30 ℃. **D)** Total RNA from haploid strains harboring the wild-type or *sen1-W1166X* alleles was resolved on denaturing agarose gels and detected by the *SNR47* probe. The amplicons used as the probe are indicated by the bars above the gene map, and the arrows mark the transcription orientation. The numbers to the left of the blots represent the length in nucleotides of the RNA marker bands. The positions of the *SNR47* and *SNR47-YDR042C* read-through transcript (*SNR47*-extend) bands are indicated. Ethidium bromide stain of the portion of the filter containing the large and small ribosomal RNAs is shown below the blots. **E)** Copper resistance assay for terminator function. Strains containing the indicated *SEN1* alleles and the *SEN1* attenuator reporter construct were serially diluted eight-fold, spotted on synthetic complete medium with no or 0.1 mM copper sulfate and incubated at 30 ℃.


A
16°C
23°C
30°C
37°C
wild-type
E1597K
W1166S
W1166X
B
46alpha
DAB206
1089-2231
1-1907
E1597K
W1166X
300 kDa
250 kDa
180 kDa
130 kDa
C
wild-type
E1597K
W1166X
CYC1
wild-type
E1597K
W1166X
SNR47
0 mM Cu
0.1 mM Cu
D
SNR47 probe
SNR47
YDR042C
2000
1500
1000
500
200
SNR47 extend
SNR47
rRNA
WT
W1166X
E
SEN1 attenuator
WT
E1597K
W1166X
0 mM Cu
0.1 mM Cu
30°C

APPENDIX 2: OBSERVING SEN1 IN LIVE YEAST CELLS USING TIME-LAPSE FLUORESCENCE MICROSCOPY

A2.1 Introduction

In Chapter 3, I discussed my live cell imaging experiments of using the plasmid-borne Sen1-GFP constructs to identify the correlation between nuclear localization of the Sen1 helicase domain and yeast cell viability. Yeast cells that contain the Sen1-GFP constructs provide a means to further explore unknown features of Sen1, such as changes and potential functions of Sen1's localization and abundance during the cell cycle. However, there was no available approach for the time course study of Sen1 in live cells. In this section, I will describe my work to establish the system that allows observing time lapse of Sen1 in live yeast cells and the results. I will also show live cell images of GFP-labeled full-length Sen1 and Sen1(964-1907) that includes the N-terminal NLS (Section 3.3) taken by N-SIM super resolution microscope (Nikon).

A2.2 Sen1 is localized inside the nucleus during mitotic cell division

Using the strain XCY361 that contains the pRS315-SEN1-GFP construct, I investigated the distribution of full-length, wild-type Sen1 in yeast cells. This experiment was conducted using the ONIX microfluidic platform, and cells were imaged using Zeiss Axiovert 200M inverted microscope (see Section 2.4). Images were taken at 7 time points (0, 30, 50, 70, 90, 110, and 130 minutes).

This system allowed me to observe the timelapse of GFP-labeled Sen1 in live yeast cells. As shown in Figure A2.1, I can observe ~300 cells in one imaging field, and over 60% of the cells are expressing GFP at a detectable level. The number of fluorescent cells is sufficient for finding individual examples repenting different phases of mitosis. I saw GFP signal in cells at all mitotic phases, indicating that Sen1 is expressed continuously. The time-lapse images confirmed that Sen1 is predominant localized inside the nucleus throughout mitosis. The fluorescent microscopy result revealed that Sen1 in the mother cell migrates through the bud neck into the

daughter cell during anaphase.

For the super resolution imaging, XCY361 cells transformed with GFP-tagged full-length Sen1 or Sen1(964-1907) in pRS315 were prepared using the protocol described in Section 2.4 Part 1. The N-SIM super resolution microscope is provided by Biochemistry Optical Core, University of Wisconsin-Madison. Localization of the Sen1-GFP constructs was visualized by the GFP channel. Shown in Figure A2.2, both full-length and the truncated Sen1 constructs are enriched in the nucleus. An "anaphase bridge" shaped structure was observed in some cells that carry the full-length Sen1-GFP. For the Sen1(964-1907)-GFP construct, I observed the crescent-shaped distribution which was found with the Sen1(1004-1907)GFP construct (see Section 3.3).

The success of using the ONIX system to observe the timelapse of Sen1 enables us to further investigate the properties of this protein. In order to characterize the potential dynamic changes of Sen1 expression during mitosis, the experiment need to be extended to a longer time course, so that cells that have gone through several rounds of the cell cycle can be identified for observing Sen1 in them. In addition to the wild-type protein, the properties of its mutants could be investigated in live cells, which would suggest the defects caused by the mutations. The Sen1-GPF constructs could also be transformed into diploid cells to characterize the functions of Sen1 during meiosis.

FIGURE A2.1 Observing dynamics of Sen1-GFP in live yeast cells by time-lapse fluorescent microscopy. Each micrograph is a merged picture of fluorescence of the Sen1-GPF construct (green) and phase contrast image of cells (gray) taken at the same time point. The time points are indicated above the micrographs. The overview at 0 minute is shown in the upper left corner. It is zoomed in to the boxed area and shown by the micrograph on its right side. The time course images of the boxed area are shown in chronological order.

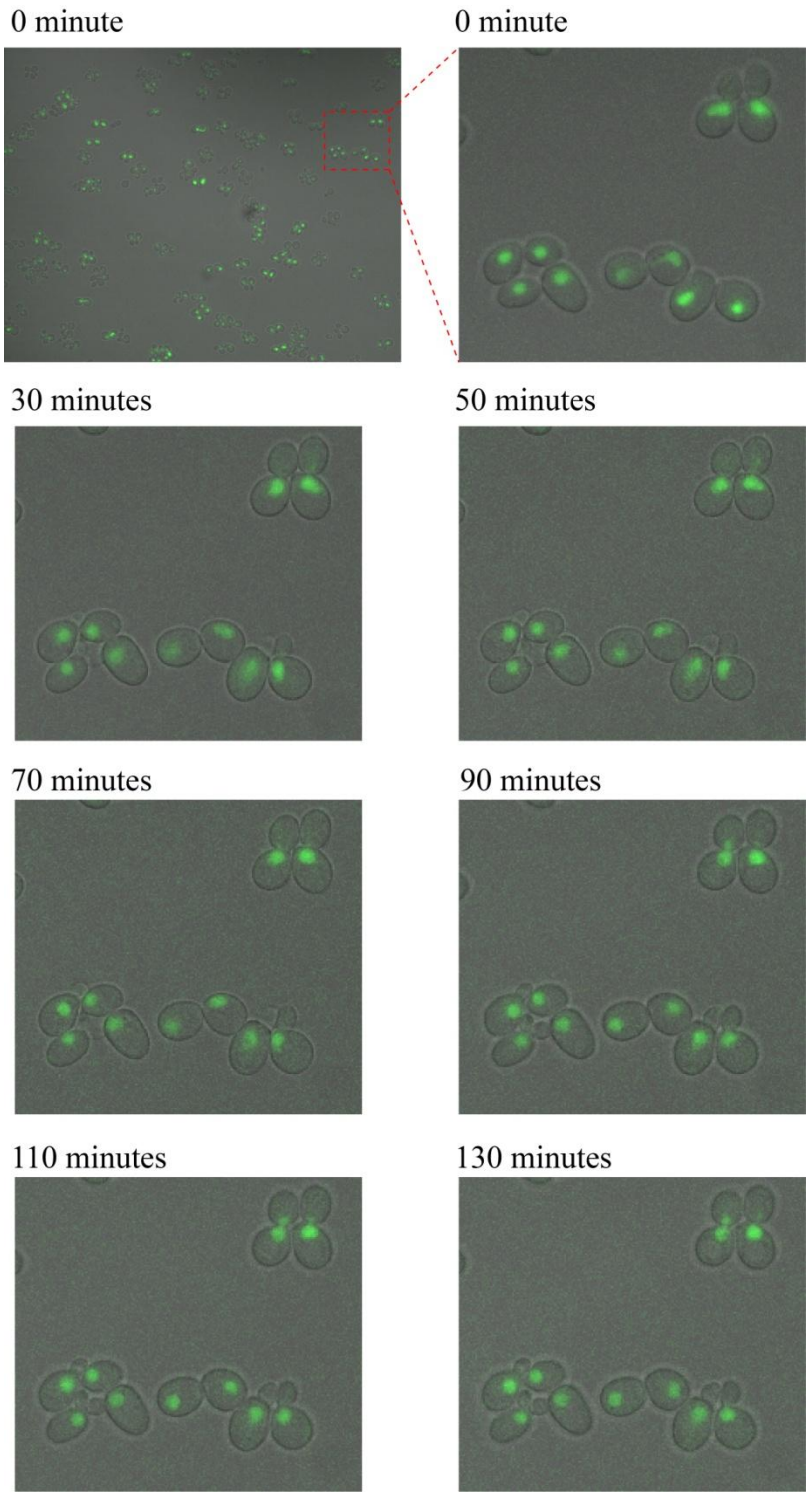
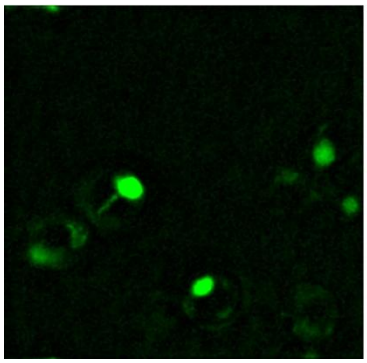
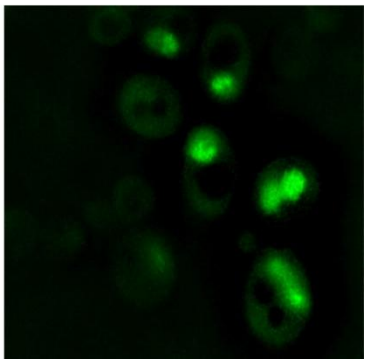

0 minute
0 minute
30 minutes
50 minutes
70 minutes
90 minutes
110 minutes
130 minutes

FIGURE A2.2 Observing localization of the Sen1-GFP constructs in live yeast cells by the N-SIM super resolution microscope. The micrograph on the left shows the cells carrying full-length Sen1-GFP. The micrograph on the right shows the distribution of Sen1(964-1907)-GFP in live cells.

FL-Sen1-GFP



Sen1(964-1907)-GFP



APPENDIX 3: CHARACTERIZING THE EFFECTS CAUSED BY A PREMATURE STOP CODON MUTATION OF NAB3

A3.1 Introduction

In 2002, Eric Steinmetz performed a genome-wide selection for spontaneous *trans*-acting read-through mutations of the *SNR13* terminator using the method described in STEINMETZ *et al.* 2006a. In this selection, he isolated 15 mutant strains that are candidates for non-*SEN1* terminator read-through mutations. One of the terminator read-through mutants, named **13α.6.1**, exhibited the cold-sensitive growth phenotype, and was identified to carry a premature stop codon mutation (CAA to TAA) in Codon 780 of *NAB3* by partially sequencing the *NAB3* gene. Since the *nab3-11* mutant exhibited termination read-through of *SNR13* (Figure 4.6A), Eric Steinmetz hypothesized that the *SNR13* terminator read-through effect is caused by the premature stop codon mutation of *NAB3*. However, this hypothesis has not been validated yet, and Eric Steinmetz did not preserve the 13α.6.1 strain. In agreement with Eric Steinmetz's finding, Loya *et al.* later found that the C-terminal tail of Nab3 where this premature stop codon mutation resides is important for Nab3 self-assembly, and another two premature stop codon mutations within this region could cause *IMD2* terminator read-through (LOYA *et al.* 2012; LOYA *et al.* 2013b). In this section, I will describe my work of verifying Eric Steinmetz's *NAB3* mutation, creating a strain that harbors a plasmid-borne allele of this *NAB3* mutant, and characterizing effects caused by this mutation.

A3.2 Materials and methods

Plasmids

To make the pRS313-nab3-Q778X construct, I used the oligos NAB3(-500)F_BamHI and NAB3(+2609)R_SalI to amplify *NAB3* from the genomic DNA of the 13α.6.1 strain with a *BamH*I and an *Sal*I sites introduced at each end. The PCR product was digested with *BamH*I and *Sal*I, and ligated to *BamH*I/*Sal*I-cut pRS313. To make the pRS313-nab3-1-777 construct, I used

pRS313-nab3-Q778X as the template and deleted Codons 778 to 802 by the QuikChange procedure, so Codon 777 is followed by the *NAB3* stop codon in pRS313-nab3-1-777.

Strains

The strain JNK004 (*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS316-*NAB3*]) was made by Jason Kuehner. The strain XCY373 (*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nrd1-V368G nab3::KanMX* [pRS316-*NAB3*]) is described in Section 2.2. Plasmid shuffle experiments were performed by plating cells on the medium containing 5-FOA at 30 ℃ to select for cells that had lost the URA3-marked wild-type *NAB3* allele.

A3.3 The 13α.6.1 strain carries a premature stop codon mutation *nab3-Q778X*

To confirm the premature stop codon is the only mutation in *NAB3* of the 13α.6.1 strain, I used oligos NAB3(-571)F and NAB3(+2659)R, PCR-amplified the entire *NAB3* gene from the genomic DNA of the 13α.6.1 strain prepared by Eric Steinmetz, and sequenced across the PCR product. According to the sequencing result, there is only one mutation in *NAB3* of the 13α.6.1 strain, which is the premature stop mutation **Q778X** (Codon ACC to TAA).

A3.4 The *nab3-Q778X* mutant allele confers slow growth phenotype

To characterize the effects caused by the *NAB3* mutation, I transformed pRS313-nab3-Q778X into JNK004 and performed plasmid shuffling by 5-FOA, which rendered a viable haploid strain named XCY389 (*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS313-*nab3-Q778X*]).

In addition to truncation of the C-terminal region after the premature stop codon, the translation read-through effect of the premature stop codon may happen on *NAB3 Q778X*. To compare the effects caused by the *NAB3* premature stop codon mutation and the *NAB3* truncation that has codons after the premature stop codon deleted, I generated a C-terminal truncation

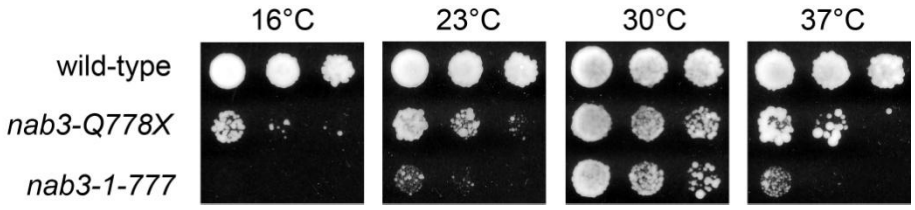
mutant by transforming pRS313-nab3-1-777 into JNK004 followed by plasmid shuffling by 5-FOA, which rendered a viable haploid strain named XCY387 (*MATα cup1Δ ura3 his3 trp1 lys2 ade2 leu2 nab3::KanMX* [pRS313-*nab3-1-777*]).

Meanwhile, I want to assess the synthetic effects caused by the Nrd1 mutant (*nrd1-5*) and Nab3 mutant (*nab3-Q778X or nab3-1-777*). After transforming pRS313-nab3-Q778X or pRS313-nab3-1-777 into XCY373 carrying *nrd1-5*, I performed plasmid shuffle experiments, but neither of the constructs rendered a viable, haploid *NAB3* mutant strain at either 23 ℃ or 30 ℃.

Next, I conducted the temperature sensitivity assay to identify the growth phenotypes of the two *NAB3* mutants. Both the premature stop codon mutant (*nab3-Q778X)* and the truncation mutant (*nab3-1-777*) exhibited slow growth phenotypes at 30 ℃, and the defects were more severe at lower or higher temperatures (16 ℃, 23 ℃ and 37 ℃) (Figure A3.1). Moreover, the truncation mutant showed more severe growth defects than the premature stop codon mutant. These observations are consistent with the hypothesis that translation read-through occurs in the premature stop codon mutant and the read-through product could partially compensate for the defects caused by the C-terminal truncated Nab3.

Collectively, I confirmed the premature stop codon mutation in a previously isolated *SNR13* terminator read-through mutant, which is Q778X. Both the premature stop codon mutant (*nab3-Q778X*) and the truncation mutant (*nab3-1-777*) were generated in a plasmid-based yeast genetic system, and exhibited growth defects at all the tested temperatures. Translation read-through of the premature stop codon may occur and partially compensate for the defects caused by the C-terminal truncation of Nab3. To test the effects of these mutants on Sen1-dependent termination, the terminator-CUP1 reporter constructs could be transformed into these mutants, and the terminator read-through level can be measure by the copper resistance assay.

FIGURE A3.1 Characterization of the *NAB3* mutants. Growth phenotypes at 16, 23, 30 and 37 ℃ of the haploid wild-type and mutants harboring *nab3-Q778X* or *nab3-1-777* after eight-fold serial dilution on YEPD medium.


16°C
23°C
30°C
37°C
wild-type
nab3-Q778X
nab3-1-777

REFERENCES

Abou Elela, S., H. Igel and M. Ares, 1996 RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. Cell 85**:** 115-124.

Allmang, C., J. Kufel, G. Chanfreau, P. Mitchell, E. Petfalski *et al.*, 1999 Functions of the exosome in rRNA, snoRNA and snRNA synthesis. Embo Journal 18**:** 5399-5410.

Alzu, A., R. Bermejo, M. Begnis, C. Lucca, D. Piccini *et al.*, 2012 Senataxin Associates with Replication Forks to Protect Fork Integrity across RNA-Polymerase-II-Transcribed Genes. Cell 151**:** 835-846.

Anheim, M., B. Monga, M. Fleury, P. Charles, C. Barbot *et al.*, 2009 Ataxia with oculomotor apraxia type 2: clinical, biological and genotype/phenotype correlation study of a cohort of 90 patients. Brain 132**:** 2688-2698.

Arigo, J. T., K. L. Carroll, J. M. Ames and J. L. Corden, 2006a Regulation of yeast *NRD1* expression by premature transcription termination. Molecular Cell 21**:** 641-651.

Arigo, J. T., D. E. Eyler, K. L. Carroll and J. L. Corden, 2006b Termination of cryptic unstable transcripts is directed by yeast RNA-Binding proteins Nrd1 and Nab3. Molecular Cell 23**:** 841-851.

Arndt, K. M., and D. Reines, 2015 Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. Annu Rev Biochem DOI: 10.1146/annurev-biochem-060614-034457.

Arning, L., J. T. Epplen, E. Rahikkala, C. Hendrich, A. C. Ludolph *et al.*, 2013 The *SETX* missense variation spectrum as evaluated in patients with ALS4-like motor neuron diseases. Neurogenetics 14**:** 53-61.

Asaka, T., H. Yokoji, J. Ito, K. Yamaguchi and A. Matsushima, 2006 Autosomal recessive ataxia with peripheral neuropathy and elevated AFP: novel mutations in *SETX*. Neurology 66**:** 1580-1581.

Bacikova, V., J. Pasulka, K. Kubicek and R. Stefl, 2014 Structure and semi-sequence-specific RNA binding of Nrd1. Nucleic Acids Research 42**:** 8024-8038.

Ballarino, M., M. Morlando, F. Pagano, A. Fatica and I. Bozzoni, 2005 The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in *Saccharomyces cerevisiae*. Mol Cell Biol 25**:** 5396-5403.

Bassuk, A. G., Y. Z. Chen, S. D. Batish, N. Nagan, P. Opal *et al.*, 2007 In cis autosomal dominant mutation of Senataxin associated with tremor/ataxia syndrome. Neurogenetics 8**:** 45-49.

Bernard, V., S. Stricker, F. Kreuz, M. Minnerop, G. Gillessen-Kaesbach *et al.*, 2008 Ataxia with oculomotor apraxia type 2: novel mutations in six patients with juvenile age of onset and elevated serum alpha-fetoprotein. Neuropediatrics 39**:** 347-350.

Boeke, J. D., J. Trueheart, G. Natsoulis and G. R. Fink, 1987 5-Fluoroorotic Acid as a Selective Agent in Yeast Molecular-Genetics. Methods in Enzymology 154**:** 164-175.

Brar, G. A., M. Yassour, N. Friedman, A. Regev, N. T. Ingolia *et al.*, 2012 High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. Science 335**:** 552-557.

Breuer, U., and H. Harms, 2006 *Debaryomyces hansenii* - an extremophilic yeast with biotech nological potential. Yeast 23**:** 415-437.

Brow, D. A., 2011 Sen-sing RNA Terminators. Molecular Cell 42**:** 717-718.

Carroll, K. L., R. Ghirlando, J. M. Ames and J. L. Corden, 2007 Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA 13**:** 361-373.

Carroll, K. L., D. A. Pradhan, J. A. Granek, N. D. Clarke and J. L. Corden, 2004 Identification of *cis* elements directing termination of yeast nonpolyadenylated snoRNA transcripts. Molecular and Cellular Biology 24**:** 6241-6252.

Chakrabarti, S., U. Jayachandran, F. Bonneau, F. Fiorini, C. Basquin *et al.*, 2011 Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. Molecular Cell 41**:** 693-703.

Chanfreau, G., P. Legrain and A. Jacquier, 1998 Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. Journal of Molecular Biology 284**:** 975-988.

Chen, X., U. Muller, K. E. Sundling and D. A. Brow, 2014 *Saccharomyces cerevisiae* Sen1 as a Model for the Study of Mutations in Human Senataxin That Elicit Cerebellar Ataxia. Genetics 198**:** 577-90.

Chen, Y. Z., C. L. Bennett, H. M. Huynh, I. P. Blair, I. Puls *et al.*, 2004 DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). American Journal of Human Genetics 74**:** 1128-1135.

Chinchilla, K., J. B. Rodriguez-Molina, D. Ursic, J. S. Finkel, A. Z. Ansari *et al.*, 2012 Interactions of Sen1, Nrd1, and Nab3 with Multiple Phosphorylated Forms of the Rpb1 C-Terminal Domain in *Saccharomyces cerevisiae*. Eukaryotic Cell 11**:** 417-429.

Conrad, M. N., A. M. Dominguez and M. E. Dresser, 1997 Ndj1p, a meiotic telomere protein required for normal chromosome synapsis and segregation in yeast. Science 276**:** 1252-1255.

Conrad, N. K., S. M. Wilson, E. J. Steinmetz, M. Patturajan, D. A. Brow *et al.*, 2000 A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics 154**:** 557-571.

Corden, J. L., 1990 Tails of Rna Polymerase II. Trends in Biochemical Sciences 15**:** 383-387.

Creamer, T. J., M. M. Darby, N. Jamonnak, P. Schaughency, H. P. Hao *et al.*, 2011 Transcriptome-Wide Binding Sites for Components of the *Saccharomyces cerevisiae* Non-Poly(A) Termination Pathway: Nrd1, Nab3, and Sen1. PLoS Genet 7(10): e1002329. doi:10.1371/journal.pgen.1002329.

Criscuolo, C., L. Chessa, S. Di Giandomenico, P. Mancini, F. Sacca *et al.*, 2006 Ataxia with oculomotor apraxia type 2: a clinical, pathologic, and genetic study. Neurology 66**:** 1207-1210.

Datta, N., and A. Hohler, 2013 A new *SETX* mutation producing AOA2 in two siblings. The International journal of neuroscience 123**:** 670-673.

Davis, M. Y., C. D. Keene, P. D. Swanson, C. Sheehy and T. D. Bird, 2013 Novel mutations in ataxia telangiectasia and AOA2 associated with prolonged survival. Journal of the neurological sciences 335**:** 134-138.

DeMarini, D. J., M. Winey, D. Ursic, F. Webb and M. R. Culbertson, 1992 *SEN1*, a positive effector of tRNA-splicing endonuclease in *Saccharomyces cerevisiae*. Mol Cell Biol 12**:** 2154-2164.

Duquette, A., K. Roddier, J. McNabb-Baltar, I. Gosselin, A. St-Denis *et al.*, 2005 Mutations in senataxin responsible for Quebec cluster of ataxia with neuropathy. Ann Neurol 57**:** 408-414.

Fairman-Williams, M. E., U. P. Guenther and E. Jankowsky, 2010 SF1 and SF2 helicases: family matters. Current Opinion in Structural Biology 20**:** 313-324.

Finkel, J. S., K. Chinchilla, D. Ursic and M. R. Culbertson, 2010 Sen1p Performs Two Genetically Separable Functions in Transcription and Processing of U5 Small Nuclear RNA in *Saccharomyces cerevisiae*. Genetics 184**:** 107-118.

Fogel, B. L., E. Cho, A. Wahnich, F. Y. Gao, O. J. Becherel *et al.*, 2014 Mutation of senataxin alters disease-specific transcriptional networks in patients with ataxia with oculomotor apraxia type 2. Human Molecular Genetics 23**:** 4758-4769.

Fogel, B. L., and S. Perlman, 2006 Novel mutations in the senataxin DNA/RNA helicase domain in ataxia with oculomotor apraxia 2. Neurology 67**:** 2083-2084.

Gazulla, J., I. Benavente, I. P. Lopez-Fraile, C. Tordesillas, P. Modrego *et al.*, 2010 Sensory neuronopathy in ataxia with oculomotor apraxia type 2. Journal of the neurological sciences 298**:** 118-120.

Gesteland, R. F., and J. F. Atkins, 1996 Recoding: Dynamic reprogramming of translation. Annual Review of Biochemistry 65**:** 741-768.

Ghazal, G., J. Gagnon, P. E. Jacques, J. R. Landry, F. Robert *et al.*, 2009 Yeast RNase III Triggers Polyadenylation-Independent Transcription Termination. Molecular Cell 36**:** 99-109.

Ghrooda, S., A. Borys, E. Spriggs, M. Hegde and A. Mhanni, 2012 SETX gene novel mutations in a non-French Canadian with ataxia-oculomotor apraxia type 2. Parkinsonism & related disorders 18**:** 700-701.

Groenewald, M., H. M. Daniel, V. Robert, G. A. Poot and M. T. Smith, 2008 Polyphasic re-examination of *Debaryomyces hansenii* strains and reinstatement of *D. hansenii*, *D.fabryi* and *D.subgiobosus*. Persoonia 21**:** 17-27.

Gudipati, R. K., T. Villa, J. Boulay and D. Libri, 2008 Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nature Structural & Molecular Biology 15**:** 786-794.

Hammer, M. B., G. El Euch-Fayache, H. Nehdi, D. Saidi, A. Nasri *et al.*, 2012 Clinical and molecular findings of ataxia with oculomotor apraxia type 2 (AOA2) in 5 Tunisian families. Diagnostic molecular pathology : the American journal of surgical pathology, part B 21**:** 241-245.

Haruki, H., J. Nishikawa and U. K. Laemmli, 2008 The anchor-away technique: Rapid, conditional establishment of yeast mutant phenotypes. Molecular Cell 31**:** 925-932.

Hatchi, E., K. Skourti-Stathaki, S. Ventz, L. Pinello, A. Yen *et al.*, 2015 BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. Mol Cell 57**:** 636-647.

Hazelbaker, D. Z., S. Marquardt, W. Wlotzka and S. Buratowski, 2013 Kinetic Competition between RNA Polymerase II and Sen1-Dependent Transcription Termination. Molecular Cell 49**:** 55-66.

He, X. Y., A. U. Khan, H. L. Cheng, D. L. Pappas, M. Hampsey *et al.*, 2003 Functional interactions between the transcription and mRNA 3 ' end processing machineries mediated by Ssu72 and Sub1. Genes & Development 17**:** 1030-1042.

Hill, S. J., T. Rolland, G. Adelmant, X. F. Xia, M. S. Owen *et al.*, 2014 Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. Genes & Development 28**:** 1957-1975.

Hobor, F., R. Pergoli, K. Kubicek, D. Hrossova, V. Bacikova *et al.*, 2011 Recognition of Transcription Termination Signal by the Nuclear Polyadenylated RNA-binding (NAB) 3 Protein. Journal of Biological Chemistry 286**:** 3645-3657.

Jamonnak, N., T. J. Creamer, M. M. Darby, P. Schaughency, S. J. Wheelan *et al.*, 2011 Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. RNA 17**:** 2011-2025.

Jenks, M. H., T. W. O'Rourke and D. Reines, 2008 Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. Molecular and Cellular Biology 28**:** 3883-3893.

Kawauchi, J., H. Mischo, P. Braglia, A. Rondon and N. J. Proudfoot, 2008 Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. Genes & Development 22**:** 1082-1092.

Kessler, M. M., M. F. Henry, E. Shen, J. Zhao, S. Gross *et al.*, 1997 Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. Genes & Development 11**:** 2545-2556.

Kim, H., B. Erickson, W. Luo, D. Seward, J. H. Graber *et al.*, 2010 Gene-specific RNA polymerase II phosphorylation and the CTD code. Nature Structural & Molecular Biology 17: 1279-86.

Kim, H. D., J. Choe and Y. S. Seo, 1999 The *sen1*(+) gene of *Schizosaccharomyces pombe*, a homologue of budding yeast *SEN1*, encodes an RNA and DNA helicase. Biochemistry 38**:** 14697-14710.

Kim, M., S. H. Ahn, N. J. Krogan, J. F. Greenblatt and S. Buratowski, 2004a Transitions in RNA polymerase II elongation complexes at the 3 ' ends of genes. Embo Journal 23**:** 354-364.

Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea *et al.*, 2004b The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432**:** 517-522.

Kim, M., L. Vasiljeva, O. J. Rando, A. Zhelkovsky, C. Moore *et al.*, 2006 Distinct pathways for snoRNA and mRNA termination. Molecular Cell 24**:** 723-734.

Komarnitsky, P., E. J. Cho and S. Buratowski, 2000 Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development 14**:** 2452-2460.

Krishnamurthy, S., X. Y. He, M. Reyes-Reyes, C. Moore and M. Hampsey, 2004 Ssu72 is an RNA polymerase II CTD phosphatase. Molecular Cell 14**:** 387-394.

Kubicek, K., H. Cerna, P. Holub, J. Pasulka, D. Hrossova *et al.*, 2012 Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes & Development 26**:** 1891-1896.

Kuehner, J. N., and D. A. Brow, 2008 Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. Molecular Cell 31**:** 201-211.

Kuehner, J. N., E. L. Pearson and C. Moore, 2011 Unravelling the means to an end: RNA polymerase II transcription termination. Nature Reviews Molecular Cell Biology 12**:** 283-294.

LaCava, J., J. Houseley, C. Saveanu, E. Petfalski, E. Thompson *et al.*, 2005 RNA degradation by the exosome is promoted by a nuclear polyade nylation complex. Cell 121**:** 713-724.

Lardenois, A., Y. C. Liu, T. Walther, F. Chalmel, B. Evrard *et al.*, 2011 Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the

conserved exosome subunit Rrp6. Proceedings of the National Academy of Sciences of the United States of America 108**:** 1058-1063.

Lemmens, R., M. J. Moore, A. Al-Chalabi, R. H. Brown and W. Robberecht, 2010 RNA metabolism and the pathogenesis of motor neuron diseases. Trends in Neurosciences 33**:** 249-258.

Lesser, C. F., and C. Guthrie, 1993 Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, *CUP1*. Genetics 133**:** 851-863.

Li, Z. J., F. J. Vizeacoumar, S. Bahr, J. J. Li, J. Warringer *et al.*, 2011 Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nature Biotechnology 29**:** 361-U105.

Logan, J., E. Falckpedersen, J. E. Darnell and T. Shenk, 1987 A Poly(A) Addition Site and a Downstream Termination Region Are Required for Efficient Cessation of Transcription by RNA Polymerase-II in the Mouse Beta Maj-Globin Gene. Proceedings of the National Academy of Sciences of the United States of America 84**:** 8306-8310.

Loya, T. J., T. W. O'Rourke, N. Degtyareva and D. Reines, 2013a A Network of Interdependent Molecular Interactions Describes a Higher Order Nrd1-Nab3 Complex Involved in Yeast Transcription Termination. Journal of Biological Chemistry 288**:** 34158-34167.

Loya, T. J., T. W. O'Rourke and D. Reines, 2012 A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. Nucleic Acids Research 40**:** 7476-7491.

Loya, T. J., T. W. O'Rourke and D. Reines, 2013b Yeast Nab3 Protein Contains a Self-assembly Domain Found in Human Heterogeneous Nuclear Ribonucleoprotein-C (hnRNP-C) That Is Necessary for Transcription Termination. Journal of Biological Chemistry 288**:** 2111-2117.

Marfori, M., A. Mynott, J. J. Ellis, A. M. Mehdi, N. F. W. Saunders *et al.*, 2011 Molecular basis for specificity of nuclear import and prediction of nuclear localization. Biochim Biophys Acta 1813**:** 1562-1577.

Marquardt, S., D. Z. Hazelbaker and S. Buratowski, 2011 Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. Transcription 2**:** 145-154.

Mayer, A., M. Heidemann, M. Lidschreiber, A. Schreieck, M. Sun *et al.*, 2012 CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science 336**:** 1723-1725.

Miller, M. S., A. Rialdi, J. S. Y. Ho, M. Tilove, L. Martinez-Gil *et al.*, 2015 Senataxin suppresses the antiviral transcriptional response and controls viral biogenesis. Nature Immunology 16**:** 485-U198.

Mischo, H. E., B. Gomez-Gonzalez, P. Grzechnik, A. G. Rondon, W. Wei *et al.*, 2011 Yeast Sen1 Helicase Protects the Genome from Transcription-Associated Instability. Molecular Cell 41**:** 21-32.

Moreira, M.-C., S. Klur, M. Watanabe, A. H. Nemeth, I. Le Ber *et al.*, 2004 Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. Nature genetics 36**:** 225-227.

Mosley, A. L., S. G. Pattenden, M. Carey, S. Venkatesh, J. M. Gilmore *et al.*, 2009 Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. Molecular Cell 34**:** 168-178.

Nakagawa, T., and R. D. Kolodner, 2002 *Saccharomyces cerevisiae* Mer3 is a DNA helicase involved in meiotic crossing over. Molecular and Cellular Biology 22**:** 3281-3291.

Nakagawa, T., and H. Ogawa, 1999 The *Saccharomyces cerevisiae* MER3 gene, encoding a novel helicase-like protein, is required for crossover control in meiosis. Embo Journal 18**:** 5714-5723.

Nanetti, L., S. Cavalieri, V. Pensato, A. Erbetta, D. Pareyson *et al.*, 2013 *SETX* mutations are a frequent genetic cause of juvenile and adult onset cerebellar ataxia with neuropathy and elevated serum alpha-fetoprotein. Orphanet Journal of Rare Diseases 8**:** 123.

Nedea, E., D. Nalbant, D. Xia, N. T. Theoharis, B. Suter *et al.*, 2008 The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Molecular Cell 29**:** 577-587.

Nguyen, H. V., C. Gaillardin and C. Neuveglise, 2009 Differentiation of *Debaryomyces hansenii* and *Candida famata* by rRNA gene intergenic spacer fingerprinting and reassessment of phylogenetic relationships among *D. hansenii, C. famata, D. fabryi, C. flareri* (*=D-subglobosus*) and *D. prosopidis*: description of *D. vietnamensis sp. nov.* closely related to *D. nepalensis*. FEMS Yeast Research 9**:** 641-662.

Nicol, J. W., G. A. Helt, S. G. Blanchard, A. Raja and A. E. Loraine, 2009 The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25**:** 2730-2731.

Ozsolak, F., P. Kapranov, S. Foissac, S. W. Kim, E. Fishilevich *et al.*, 2010 Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation. Cell 143**:** 1018-1029.

Pelechano, V., W. Wei, P. Jakob and L. M. Steinmetz, 2014 Genome-wide identification of transcript start and end sites by transcript isoform sequencing. Nature Protocols 9**:** 1740-1759.

Perez-Canadillas, J. M., 2006 Grabbing the message: structural basis of mRNA 3 ' UTR recognition by Hrp1. Embo Journal 25**:** 3167-3178.

Pierce, S. E., R. W. Davis, C. Nislow and G. Giaever, 2007 Genome-wide analysis of barcoded *Saccharomyces cerevisiae* gene-deletion mutants in pooled cultures. Nature Protocols 2**:** 2958-2974.

Piluso, G., M. Dionisi, F. D. Blanco, A. Torella, S. Aurino *et al.*, 2011 Motor Chip: A Comparative Genomic Hybridization Microarray for Copy-Number Mutations in 245 Neuromuscular Disorders. Clinical Chemistry 57**:** 1584-1596.

Poorey, K., R. O. Sprouse, M. N. Wells, R. Viswanathan, S. Bekiranov *et al.*, 2010 RNA synthesis precision is regulated by preinitiation complex turnover. Genome Research 20**:** 1679-1688.

Porrua, O., F. Hobor, J. Boulay, K. Kubicek, Y. D'Aubenton-Carafa *et al.*, 2012 *in vivo* SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. Embo Journal 31**:** 3935-3948.

Porrua, O., and D. Libri, 2013 A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nature Structural & Molecular Biology 20**:** 884-91.

Porrua, O., and D. Libri, 2015 Transcription termination and the control of the transcriptome: why, where and how to stop. Nature Reviews Molecular Cell Biology 16**:** 190-202.

Rasmussen, T. P., and M. R. Culbertson, 1998 The putative nucleic acid helicase sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in *Saccharomyces cerevisiae*. Molecular and Cellular Biology 18**:** 6885-6896.

Richard, P., and J. L. Manley, 2009 Transcription termination by nuclear RNA polymerases. Genes & Development 23**:** 1247-1269.

Rondon, A. G., H. E. Mischo, J. Kawauchi and N. J. Proudfoot, 2009 Fail-Safe Transcriptional Termination for Protein-Coding Genes in *S. cerevisiae*. Molecular Cell 36**:** 88-98.

Roy, D., Z. Zhang, Z. F. Lu, C. L. Hsieh and M. R. Lieber, 2010 Competition between the RNA Transcript and the Nontemplate DNA Strand during R-Loop Formation *in vitro*: a Nick Can Serve as a Strong R-Loop Initiation Site. Molecular and Cellular Biology 30**:** 146-159.

Rudnik-Schoneborn, S., L. Arning, J. T. Epplen and K. Zerres, 2012 *SETX* gene mutation in a family diagnosed autosomal dominant proximal spinal muscular atrophy. Neuromuscular disorders : NMD 22**:** 258-262.

Saiga, T., T. Tateishi, T. Torii, N. Kawamura, Y. Nagara *et al.*, 2012 Inflammatory radiculoneuropathy in an ALS4 patient with a novel *SETX* mutation. J Neurol Neurosurg Psychiatry 83**:** 763-764.

Schaughency, P., J. Merran and J. L. Corden, 2014 Genome-Wide Mapping of Yeast RNA Polymerase II Termination. PLoS Genet 10(10): e1004632. doi:10.1371/journal.pgen.1004632.

Schiestl, R. H., and R. D. Gietz, 1989 High-Efficiency Transformation of Intact Yeast Cells Using Single Stranded Nucleic Acids as a Carrier. Current Genetics 16**:** 339-346.

Schmitt, M. E., T. A. Brown and B. L. Trumpower, 1990 A Rapid and Simple Method for Preparation of RNA from *Saccharomyces Cerevisiae*. Nucleic Acids Research 18**:** 3091-3092.

Schulz, D., B. Schwalb, A. Kiesel, C. Baejen, P. Torkler *et al.*, 2013 Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. Cell 155**:** 1075-1087.

Singh, N., Z. Ma, T. Gemmill, X. Y. Wu, H. DeFiglio *et al.*, 2009 The Ess1 Prolyl Isomerase Is Required for Transcription Termination of Small Noncoding RNAs via the Nrd1 Pathway. Molecular Cell 36: 255-266.

Skourti-Stathaki, K., and N. J. Proudfoot, 2014 A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & Development 28**:** 1384-1396.

Skourti-Stathaki, K., N. J. Proudfoot and N. Gromak, 2011 Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Molecular Cell 42**:** 794-805.

Smith, J. E., J. R. Alvarez-Dominguez, N. Kline, N. J. Huynh, S. Geisler *et al.*, 2014 Translation of Small Open Reading Frames within Unannotated RNA Transcripts in *Saccharomyces cerevisiae*. Cell Reports 7**:** 1858-1866.

Steinmetz, E. J., and D. A. Brow, 1996 Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. Mol Cell Biol 16**:** 6993-7003.

Steinmetz, E. J., and D. A. Brow, 1998 Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proceedings of the National Academy of Sciences of the United States of America 95**:** 6699-6704.

Steinmetz, E. J., and D. A. Brow, 2003 Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. Molecular and Cellular Biology 23**:** 6339-6349.

Steinmetz, E. J., N. K. Conrad, D. A. Brow and J. L. Corden, 2001 RNA-binding protein Nrd1 directs poly(A)-independent 3 '-end formation of RNA polymerase II transcripts. Nature 413**:** 327-331.

Steinmetz, E. J., S. B. H. Ng, J. P. Cloute and D. A. Brow, 2006a *cis*- and *trans*-acting determinants of transcription termination by yeast RNA polymerase II. Molecular and Cellular Biology 26**:** 2688-2696.

Steinmetz, E. J., C. L. Warren, J. N. Kuehner, B. Panbehi, A. Z. Ansari *et al.*, 2006b Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Molecular Cell 24**:** 735-746.

Stirling, P. C., Y. J. A. Chan, S. W. Minaker, M. J. Aristizabal, I. Barrett *et al.*, 2012 R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes & Development 26**:** 163-175.

Sugiyama, T., R. Sugioka-Sugiyama, K. Hada and R. Niwa, 2012 Rhn1, a Nuclear Protein, Is Required for Suppression of Meiotic mRNAs in Mitotically Dividing Fission Yeast. PLoS ONE 7(8): e42962. doi:10.1371/journal.pone.0042962.

Sun, M., B. Schwalb, D. Schulz, N. Pirkl, S. Etzold *et al.*, 2012 Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Research 22**:** 1350-1359.

Suraweera, A., O. J. Becherel, P. Chen, N. Rundle, R. Woods *et al.*, 2007 Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. Journal of Cell Biology 177**:** 969-979.

Thiebaut, M., E. Kisseleva-Romanova, M. Rougemaille, J. Boulay and D. Libri, 2006 Transcription termination and nuclear degradation of cryptic unstable transcripts: A role for the Nrd1-Nab3 pathway in genome surveillance. Molecular Cell 23**:** 853-864.

Tuck, A. C., and D. Tollervey, 2013 A Transcriptome-wide Atlas of RNP Composition Reveals Diverse Classes of mRNAs and lncRNAs. Cell 154**:** 996-1009.

Tudek, A., O. Porrua, T. Kabzinski, M. Lidschreiber, K. Kubicek *et al.*, 2014 Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. Molecular Cell 55**:** 467-481.

Ursic, D., K. Chinchilla, J. S. Finkel and M. R. Culbertson, 2004 Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. Nucleic Acids Research 32**:** 2441-2452.

Ursic, D., D. J. DeMarini and M. R. Culbertson, 1995 Inactivation of the yeast Sen1 protein affects the localization of nucleolar proteins. Molecular & General Genetics 249**:** 571-584.

Vantaggiato, C., O. Cantoni, A. Guidarelli, R. Romaniello, A. Citterio *et al.*, 2014 Novel *SETX* variants in a patient with ataxia, neuropathy, and oculomotor apraxia are associated with normal sensitivity to oxidative DNA damaging agents. Brain Dev 36(8):682-9.

Vasiljeva, L., and S. Buratowski, 2006 Nrd1 interacts with the nuclear exosorne for 3 ' processing of RNA polymerase II transcripts. Molecular Cell 21**:** 239-248.

Vasiljeva, L., M. Kim, H. Mutschler, S. Buratowski and A. Meinhart, 2008 The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nature Structural & Molecular Biology 15**:** 795-804.

West, M. L., and J. L. Corden, 1995 Construction and Analysis of Yeast Rna-Polymerase-Ii Ctd Deletion and Substitution Mutations. Genetics 140**:** 1223-1233.

West, S., N. Gromak and N. J. Proudfoot, 2004 Human 5 '-> 3 ' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432**:** 522-525.

Wilson, S. M., K. V. Datar, M. R. Paddy, J. R. Swedlow and M. S. Swanson, 1994 Characterization of Nuclear Polyadenylated Rna-Binding Proteins in *Saccharomyces cerevisiae*. Journal of Cell Biology 127**:** 1173-1184.

Winey, M., and M. R. Culbertson, 1988 Mutations affecting the tRNA-splicing endonuclease activity of *Saccharomyces cerevisiae*. Genetics 118**:** 609-617.

Wise, J. A., 1991 Preparation and analysis of low molecular weight RNAs and small ribonucleoproteins. Methods Enzymol 194**:** 405-415.

Wlotzka, W., G. Kudla, S. Granneman and D. Tollervey, 2011 The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. Embo Journal 30**:** 2982-2982.

Wu, H. Y., and S. M. Burgess, 2006 Ndj1, a telomere-associated protein, promotes meiotic recombination in budding yeast. Molecular and Cellular Biology 26**:** 3683-3694.

Wyers, F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour *et al.*, 2005 Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121**:** 725-737.

Yang, P. K., C. Hoareau, C. Froment, B. Monsarrat, Y. Henry *et al.*, 2005 Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. Mol Cell Biol 25**:** 3295-3304.

Zhang, D. W., A. L. Mosley, S. R. Ramisetty, J. B. Rodriguez-Molina, M. P. Washburn *et al.*, 2012 Ssu72 Phosphatase-dependent Erasure of Phospho-Ser7 Marks on the RNA Polymerase II C-terminal Domain Is Essential for Viability and Transcription Termination. Journal of Biological Chemistry 287**:** 8541-8551.