

**CHARACTERIZING THE ROLE OF THE DEAD-BOX PROTEIN DBP2 IN
RNA STRUCTURE REMODELING AND PRE-MRNA PROCESSING**

by

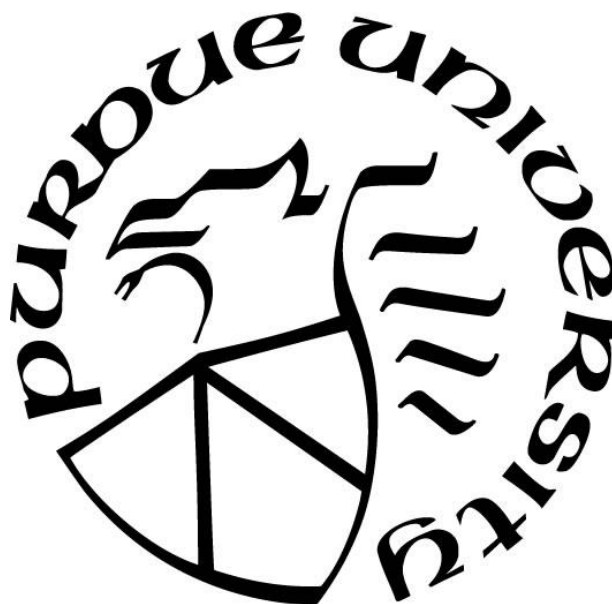
Yu-Hsuan Lai

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Biochemistry

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Elizabeth Tran, Chair

Department of Biochemistry

Dr. Barbara Golden

Department of Biochemistry

Dr. Frederick Gimble

Department of Biochemistry

Dr. Pete Pascuzzi

Libraries and School of Information Studies

Approved by:

Dr. Andrew Mesecar

Head of the Graduate Program

ACKNOWLEDGMENTS

I would like to thank everyone in the Tran lab for all the support during these years. I appreciate the help and advice provided by my advisor, Dr. Tran. I thank her for often being available for discussions and very responsive to questions. I also appreciate her giving me lots of opportunities to learn and grow, including grant writing, paper reviewing, conference presentation, and interdisciplinary collaboration. Besides, I appreciate the help from former graduate students, Kit Ma, Siwen Wang, and Cindy Xing. They were all very generous in giving constructive feedbacks on all aspects of my graduate school life, and have been very supportive whenever I reach out, even after they graduated. I also had a lot of fun with Kit, Siwen, Cindy, and Sara, and I appreciate all the little fun moments we shared.

I would also like to thank my thesis committee members: Dr. Golden, Dr. Gimble, and Dr. Pascuzzi. I am very fortunate to have them on my committee. Dr. Golden has been a very supportive and understanding mentor since my first year in grad school, and talking to her has always been refreshing and empowering. Dr. Gimble has been a great model of an effective teacher, and I learn a lot from him – including solid knowledge, logical thinking, and teaching skills. Dr. Pascuzzi is always very patient and approachable when I have bioinformatics questions, and he always provides great resource for solving problems. In addition to my committee members, I would like to thank Dr. Nadia Atallah. Nadia has been very supportive for my professional development, not only providing me help for bioinformatics analysis, but also giving me constructive advice for future career and life in general. I am very grateful for having such a great mentor and friend in my life.

I thank everyone that I met in the Department of Biochemistry and Purdue University. I sincerely appreciate all the support, help, care, or pain, difficulty, and frustration throughout these years. All these make me learn and grow to be smarter and stronger than I was before.

Finally, I would like to thank my dearest family and friends for their unconditional love.

2.5	References.....	57
CHAPTER 3. UNPUBLISHED RESULTS AND FUTURE WORK		120
3.1	Dbp2 binding is correlated with RNAPII pausing at the 5' end during transcription of protein-coding genes.....	120
3.2	Dbp2 binding at 5' end of mRNAs may be involved in premature termination by NNS120	
3.3	Future work for study of Dbp2 in premature termination and promoter-proximal RNAPII pausing	121
3.3.1	To Detect Prematurely Terminated Transcripts by Sequencing of Nascent RNAs in the Presence and Absence of <i>DBP2</i>	121
3.3.2	To Determine How Dbp2 is Recruited to the 5' ends of mRNAs	121
3.3.2.1	<i>Identification of sequence motifs enriched in Dbp2 binding regions at 5' ends of mRNAs</i>	121
3.3.2.2	<i>To test if the recruitment of Dbp2 is dependent on post-translational modifications in the RNAPII C-terminal domain</i>	122
3.4	Loss of Dbp2 Results in Accumulation of Unspliced Pre-mRNAs	122
3.5	Dbp2 Enzymatic Activity is Necessary for Efficient Splicing	124
3.6	Future Plans for Study of Dbp2 in Splicing.....	124
3.6.1	To Characterize snRNA Expression in <i>dbp2Δ</i> Compared to WT	124
3.6.2	To Compute the Folding Energy of Dbp2-Bound Introns Comparing to Introns Without Detectable Dbp2 Binding	124
3.6.3	To test Dbp2-dependent splicing of specific genes using the <i>CUPI</i> reporter system... ..	125
3.7	Perspectives.....	125
3.8	References.....	126
DECLARATION OF COLLABORATIVE WORK		136
VITA.....		137

LIST OF TABLES

Table 1.1. Genome-wide methods for detection of DEAD-box protein binding on nucleic acids	32
Table 2.1. Yeast strains. Strains are all isogenic and correspond to the BY4741 or S288C background.....	83
Table 2.2. Plasmids used in this study. Plasmids were used for endogenous, 3X-FLAG tagging or as the termination reporter (Figure 2.11).	84
Table 2.3. Primers for strain construction and cloning. Primers were used for endogenous, 3X-FLAG tagging or for reporter construction.....	85
Table 2.4. Sequences used for iCLIP-seq. Bolded letters are barcodes for different libraries.	86
Table 2.5. Oligonucleotides used for ChIP. Oligos correspond to Figure 2.3 & 2.6.....	87
Table 2.6. Oligonucleotides and PCR primers for Structure-Seq. Bolded letters are barcodes for multiple libraries.	89
Table 2.7. The list of read-through transcripts in <i>dbp2Δ</i> . The ratio of 3' extended / total transcripts is also listed for both wild type and <i>dbp2Δ</i>	90
Table 2.8. The contingency table for a Fisher's exact test of the correlation between the list of transcripts bound by Dbp2 at the 3' end and transcripts with read-through defects in <i>dbp2Δ</i> (related to Figure 2.5).....	113
Table 2.9. The list of transcripts with significant <i>DBP2</i> -dependent structural changes.....	114
Table 2.10. The contingency table for a Fisher's exact test of the correlation between the list of Dbp2-bound transcripts and transcripts with significant <i>DBP2</i> -dependent structural changes..	118
Table 2.11. The contingency table for a Fisher's exact test of the correlation between the list of transcripts with read-through defects in <i>dbp2Δ</i> and transcripts with significant <i>DBP2</i> -dependent structural changes (related to Figure 2.10)	119

LIST OF FIGURES

Figure 1.1. Potential scenarios for (A) decrease and (B) increase in RNA structural probing signals in DEAD-box protein mutant strains as compared to wild-type cells.	30
Figure 1.2. Genome-wide techniques for studying cellular functions of DEAD-box proteins. ...	31
Figure 2.1. Assessment of reproducibility of Dbp2-binding sites as determined by iCLIP-seq across three biological replicates (R1, R2, and R3).	68
Figure 2.2. Dbp2 promotes transcription termination and processing at snoRNA genes.	69
Figure 2.3. Dbp2 binding in protein-coding transcripts correlates with Nrd1 and Nab3 binding sites and shares similar RNA sequence motifs.	71
Figure 2.4. Dbp2 binding at the 3' ends of protein-coding transcripts does not correlate with Rna15 component of the cleavage and polyadenylation complex (CPF).	72
Figure 2.5. Loss of <i>DBP2</i> causes termination defects at a subset of protein-coding genes.	73
Figure 2.6. <i>DBP2</i> -dependent termination of protein-coding genes correlates with efficient <i>DBP2</i> -dependent recruitment of Nrd1 within the gene ORF and 3'UTR.	75
Figure 2.7. The frequency of RT stops at A, U, C, and G in each replicate of DMS-treated (+) and untreated (-) samples.	77
Figure 2.8. Structure-seq reveals <i>DBP2</i> -dependent RNA structural changes in protein-coding genes.	78
Figure 2.9. DMS reactivity (A, C) and Dbp2 binding (B, D) profiles of the two snoRNAs with a read-through defect in <i>dbp2Δ</i>	80
Figure 2.10. The presence of <i>DBP2</i> -dependent structural changes in 3' UTRs correlates with a requirement for <i>DBP2</i> in transcriptional termination.	81
Figure 2.11. Secondary structure stability correlates with the requirement for <i>DBP2</i> in termination of protein-coding genes.	82
Figure 3.1. Dbp2 binding correlates with RNAPII pausing at 5' ends of protein-coding genes.	131
Figure 3.2. Dbp2 and Nrd1 preferentially binds to the 5' ends of mRNAs up-regulated in <i>dbp2Δ</i> , and display diminished 3' binding compared to the other mRNAs.	132
Figure 3.3. Dbp2 promotes efficient pre-mRNA splicing.	133
Figure 3.4. The enzymatic activity of Dbp2 is required for efficient pre-mRNA splicing.	134
Figure 3.5. Schematic representation of the splicing reporter.	135

ABSTRACT

Author: Lai, Yu-Hsuan. PhD

Institution: Purdue University

Degree Received: May 2019

Title: Characterizing the Role of the DEAD-box Protein Dbp2 in RNA Structure Remodeling and Pre-mRNA Processing

Committee Chair: Elizabeth Tran

RNA helicases are found in all kingdoms of life, functioning in all aspects of RNA biology mainly through modulating structures of RNA and ribonucleoprotein (RNP) complex. RNA structures have fundamental impacts on steps in gene expression, including transcription, pre-mRNA processing, and translation. However, the precise roles and regulatory mechanisms of RNA structures in co- and post-transcriptional processes remain elusive. By probing genome-wide RNA structures *in vivo*, a recent study suggested that ATP-dependent factors, such as RNA helicases, maintain the actively unfolded state of RNAs. Among all RNA helicases, DEAD-box proteins form the largest family in eukaryotes, and have been shown to remodel RNA/RNP structures both *in vitro* and *in vivo*. Nevertheless, for the majority of these enzymes, it is largely unclear what RNAs are targeted and where they modulate RNA/RNP structures to regulate co-transcriptional processes. To fill the gap, my research focused on identification of the RNAs and structures targeted by the DEAD-box protein Dbp2 in *S. cerevisiae* to uncover the cellular processes that Dbp2 is involved in.

My studies revealed a role of Dbp2 in transcriptional termination. Dbp2 binds to ~34% of yeast mRNAs and all snoRNAs, and loss of *DBP2* leads to a termination defect as evidenced by RNA polymerase II (RNAPII) accumulation at 3' ends of these genes. In addition, the binding pattern of Dbp2 in mRNAs is highly similar to Nrd1 and Nab3 in the Nrd1-Nab3-Sen1 (NNS) termination complex, and deletion of *DBP2* leads to reduced recruitment of Nrd1 to its target genomic loci. In Dbp2 and NNS targeted 3' UTRs, RNA structural changes resulted from *DBP2* deletion also overlap polyadenylation elements and correlate with inefficient termination, and loss of stable structure in the 3' UTR bypasses the requirement for Dbp2. These findings lead to a model that Dbp2 promotes efficient termination of transcription through RNA structure remodeling.

Interestingly, my research also revealed the requirement of *DBP2* for efficient splicing, as loss of *DBP2* leads to accumulation of unspliced pre-mRNAs. Moreover, this function is dependent on the helicase activity of Dbp2. Further studies are needed to characterize the molecular mechanism of how Dbp2 facilitates splicing in cells. Overall, my research demonstrated that DEAD-box RNA helicases remodel mRNA structure *in vivo* and that structural alteration can be essential for proper gene expression.

CHAPTER 1. INTRODUCTION

1.1 Biological relevance of RNA structures

RNA molecules can form complex secondary and tertiary structures in cells. RNA structures have been shown to impact multiple steps in RNA metabolism, from transcription to translation and decay of RNAs (1, 2). For example, formation of secondary structure near splice sites of pre-mRNAs has been reported to prevent the recognition of the splicing signals by small nuclear RNAs (snRNAs) or small nuclear ribonucleoproteins (snRNPs), thereby inhibiting splicing (3–6). Folding of introns, on the other hand, can promote splicing by bringing two splice sites to close proximity (7–9). Moreover, in translation, RNA structures near translational start sites hamper initiation (10–13). Several recent genome-wide studies indicated negative correlation between the level of RNA structures around translational start site and the efficiency of translation (14–17). In the coding region, RNA structures also modulate the level of protein expression, presumably causing ribosome pausing and facilitating co-translational protein folding (18–20). Furthermore, folding of RNAs contributes to RNA stability, as RNA structures at mRNA 3' ends inhibit degradation by exosome (21).

RNA structures also have great impact on RNP assembly. Folding of RNA molecules influences the recognition by RNA-binding proteins (RBPs). Specifically, formation of RNA structures can hinder the assembly of mRNA-protein complexes since many mRBPs usually target single-stranded RNAs (22, 23). Some RNA structural motifs have also been shown to be recognized by specific RBPs and dictate the localization of mRNAs (24, 25). As a consequence, the metabolic fate of a transcript can be affected by RNA structures that regulate protein binding on the RNA. Thus, the dynamic folding of RNA transcripts adds another layer of regulation to gene expression. However, the specific effects and regulators of RNA structures in eukaryotic co- and post-transcriptional processes remain largely unknown.

One RNA molecule can often be folded into multiple thermodynamically favored conformations. However, for the RNA to perform its function properly, formation of other non-functional, yet stable, structures have to be prevented or resolved. Potential candidates for solving this RNA folding problem *in vivo* are RNA chaperones, including RNA helicases, which can bind and remodel RNAs (26, 27). By probing genome-wide RNA structures *in vivo*, a recent study also

suggested that ATP-dependent factors, such as RNA helicases, maintain the actively unfolded state of RNAs (28). Therefore, to better understand the biological function of RNA structures and their regulators, identification of the enzymatic targets of RNA helicases becomes the next critical step in the field.

1.2 Biological activities of DEAD-box helicases

RNA helicases are found in all kingdoms of life, playing central roles in all aspects of RNA metabolism (29). They have a conserved helicase core which is responsible for ATP binding, hydrolysis, and RNA binding (30). Among them, DEAD-box proteins constitute the largest RNA helicase family, with 4 members in *E. coli*, 26 in *Saccharomyces cerevisiae*, and 37 in human (31). They are characterized by the Asp-Glu-Ala-Asp (D-E-A-D) motif in the helicase core (32). Some of them have additional N-terminal or C-terminal domains, contributing to specific functions of different DEAD-box proteins (33). Many DEAD-box helicases have been shown to be involved in steps of gene expression, including transcription, splicing, export, translation, and RNA decay (30). Importantly, several human DEAD-box proteins have also been implicated in a variety of diseases and cancer development (34, 35).

Most DEAD-box helicases have an ATP-dependent RNA-unwinding activity *in vitro*. Some of them also display an RNA annealing activity that does not require ATP (36). Although non-processive, this duplex unwinding activity has been demonstrated to be functional in several RNA processing steps. For example, Mss116 in *S. cerevisiae* serves as an RNA chaperone, assisting the folding of functional group I and II introns by disrupting misfolded structures (37–39). DEAD-box helicases, including DDX5 and DDX17, have also been reported to unwind secondary RNA structures and thereby regulate alternative splicing and/or miRNA processing (40–43). In addition, the RNA remodeling activity is critical for translation. Particularly, several DEAD-box helicases have been found to resolve structures in 5' UTR to facilitate the initiation of translation and ribosome scanning (44–46). These examples elucidate that the RNA remodeling activity plays a critical role in gene expression, yet the enzymatic targets of the majority of DEAD-box family remain uncharacterized.

DEAD-box proteins bind RNAs through contact with the phosphate backbone. Therefore, without strong interaction with nucleotide bases, these helicases generally do not show sequence specificity *in vitro* (47). *In vivo*, however, some DEAD-box proteins have been found to bind

specific RNA sequences or structures (48, 49). This specificity is often conferred by other interacting proteins or cofactors. For example, an RNA-binding protein, Rrp5 in *S. cerevisiae*, physically interacts with Rok1, a DEAD-box helicase, and enhances the RNA binding and annealing activity of Rok1 on specific ribosomal RNA substrates (49). Another example is the enhancement of eIF4A helicase activity by an RNA-binding protein, eIF4G (48). The interaction between eIF4G and the N-terminal domain of eIF4A stimulates the ATPase activity, promoting initiation of translation (50). Although DEAD-box proteins usually function in specific steps of RNA metabolism, only a few factors that control the substrate specificity of DEAD-box helicases *in vivo* have been reported. Identifying the specificity and potential cofactors is obviously needed to further understand cellular functions and regulatory mechanisms of this helicase family.

Besides RNA structure remodeling, DEAD-box proteins also modulate ribonucleoprotein (RNP) complex assembly (47). For instance, binding of the DEAD-proteins eIF4AIII primes the formation of the exon junction complex (EJC) and serves as an RNA clamp (51). This activity is also regulated by other factors, as the interaction between eIF4AIII and components in EJC enhances the RNA binding affinity of eIF4AIII (51). On the other hand, Dbp5 and Ded1 in *S. cerevisiae* have been shown to displace certain protein complex from RNA by an unclear mechanism (52, 53). Although the clamping and RNP remodeling activity of most DEAD-box members have not been analyzed, it is likely that these helicases utilize multiple biochemical activities during RNA metabolism (36).

1.3 Genome-wide approaches for identification of DEAD-box helicase binding targets

For the majority of DEAD-box proteins, the precise genomic loci, RNAs, and regions in transcripts targeted by them remain unclear. Although many of them have been extensively studied *in vitro* for characterization of biochemical properties, knowledge about their biological and molecular functions inside cells is still very limited to date. Much effort has been made to study the action of DEAD-box helicases on individual RNA transcripts, providing potential mechanisms of their role in gene regulation. However, it is elusive whether the activity of DEAD-box proteins identified in gene-specific studies also represents the mechanisms of how these enzymes regulate other transcripts. Moreover, the transcript-, sequence-, or structure-specificity of DEAD-box helicases in cells is difficult to define in studies with only individual transcripts. The advance of sequencing technology in the past decade enables research at a genome-wide scale, including

identification of targets of DEAD-box proteins. Several high-throughput methods have been developed and applied to the study of cellular functions of DEAD-box helicase. These results collectively advance the understanding of their cellular roles by providing comprehensive pictures of genes targeted and impacted. In the following sections, high-throughput techniques used to identify nucleic acid targets of DEAD-box proteins will be reviewed, including their strengths and limitations.

1.3.1 Chromatin Immunoprecipitation (ChIP) combined with next-generation sequencing

ChIP has long been used to probe the association of transcription factors and RNA polymerases with chromatin. It has also been applied to detection of the co-transcriptional interaction between RBPs, including DEAD-box helicases, and chromatin (54–56). DEAD-box proteins, including DDX5, DDX17, and DDX21, have been shown to be recruited to the promoter region and act as a co-activator for transcription (57, 58), a function that can be independent of their helicase activity in some cases (59). With the genome-wide information from ChIP-seq, the co-transcriptional processes that DEAD-box proteins are involved in can be identified. For example, ChIP-seq of DDX5 (p68) reveals the coincident localization with CCCTC-binding factor (CTCF), a DNA-binding protein with insulator function in nuclear organization and gene expression (55). This co-localization and physical interaction thus elucidates a potential mechanism of how DDX5 mediates gene regulation. In another example, ChIP-seq of DDX21 provided crucial information of where DDX21 is localized in the genome, revealing transcriptional regulation of a specific subset of genes by DDX21 (54). By comparing the ChIP-seq data of DDX21 with genome-wide mapping data of RNA-DNA hybrids (R-loops), DDX21 was also found to localize to where R-loops are formed (56). As persistence of R-loops can cause DNA damages (60), this comparison revealed a role of DDX21 in the maintenance of genome stability.

Although ChIP-seq provides valuable insights of how RNA helicases function in transcription, the resolution of this approach (~200 bp) is usually not sufficient to determine precise binding sites (Table 1). In addition, the non-strand specific property of the method makes it difficult to distinguish the direction of transcription that is regulated by the target protein, especially in compact genomes. Moreover, signals from ChIP-seq may not originate directly from binding to the chromatin, but from the interaction between the helicase of interest and the transcriptional machinery. For example, human DDX5 has been shown to associate with the

chromatin using ChIP (55); however, it also interacts with RNA polymerase II (RNAPII) (61). Thus, signals derived from ChIP-seq could be contributed by many different types of interactions, including the interaction between DDX5 and DNA, RNA, or RNAPII, and this technique alone is insufficient to decipher the comprehensive role of DDX5 in transcription.

1.3.2 RNA immunoprecipitation combined with next generation sequencing (RIP-seq)

RIP has been widely utilized to identify the binding targets of RBPs. In short, the protein of interest is immunoprecipitated from cell lysates using specific antibody, assuming bound RNAs are pulled down with the protein. This can be performed in cells with or without formaldehyde crosslinking. After purification of the bound RNAs, specific sequences can be analyzed by reverse transcription coupled with quantitative polymerase chain reaction (RT-qPCR). With the advance in the sequencing technology, the extracted RNAs can also be analyzed at a transcriptome-wide scale. This facilitates our understanding of RBP functions by studying the targeted RNA species and regions.

RIP-seq has been applied to studies of several DEAD-box proteins, including eukaryotic initiation factor 4A (eIF4A), DDX21, and DDX5 in human, DDX39ab in zebrafish, and Dhh1 in *S. cerevisiae* (62–66). eIF4A is an ATP-dependent RNA helicase, facilitating translation initiation by unwinding of 5' UTR (67). Under the treatment of Rocaglamide A (RocA), a protein synthesis inhibitor and anti-tumor compound, RIP-seq of eIF4A-bound RNAs revealed enrichment of specific transcripts as compared to the untreated control (66). This indicates enhanced clamping of eIF4A, which then leads to inhibition of translation, and uncovers the mechanism of how RocA selectively suppresses the expression of certain genes. RIP-seq also serves as a tool for identification of important long non-coding RNAs (lncRNAs) regulated by DEAD-box proteins. Specifically, by performing RIP-seq of DDX5 (p68), lncRNA *LOC284454* was found as a binding target (63). DDX5 is implicated in multiple human diseases (34, 35). Several prior studies have shown that DDX5 associates with lncRNAs and regulates transcription and expression of genes in various cellular processes (68, 69). Characterization of its target lncRNAs would further uncover the gene network modulated by DDX5 and provide new direction for design of potential therapeutics. Additionally, by integrating multiple RIP-seq datasets of RBPs interacting with DEAD-box helicases, specific pathways that are regulated by these protein complexes could be identified from the commonly associated transcripts (64, 65). For instance, a factor overexpressed

during the progression of leukemia, amino-terminal enhancer of split (AES), was found to bind snoRNAs targeted by a DEAD-box helicase, DDX21, by comparing their RIP-seq data (64). This supports their physical interaction and further helps to uncover the role of DDX21 in leukemia progression by promoting snoRNP formation. RNA-seq data from DDX21 and AES knockdown cells also revealed their functional interaction as the majority of AES-dependent snoRNAs are misregulated in DDX21-knockdown cells.

Although RIP-seq provides useful information of bound transcripts, the resolution of this method is not satisfactory when it comes to identification of binding motifs (Table 1). Depending on the size of the RBP or RBP complex footprint, RIP-seq usually provides the information of binding region greater than 20 nucleotides. Thus, it is difficult to identify motifs that are preferably bound by RBPs like DEAD-box proteins, as they have a small footprint of 6-8 nucleotides (70). In addition, in RIP protocols without crosslinking, the capture of RNA targets relies on stable interaction between the tested protein and bound transcripts under the buffer condition used. Therefore, transient interactions, such as RNA helicases and the RNA structures being unwound, may not be detected using this approach.

1.3.3 Crosslinking and immunoprecipitation followed by next-generation sequencing (CLIP-seq)

In the past decade, CLIP-seq has emerged as a powerful tool for discovery of RBP binding sites in RNAs. The basic steps in this method include ultra violet (UV) crosslinking of the target protein and its bound RNAs followed by immunoprecipitation (IP), proteinase K digestion, RNA isolation, and reverse transcription (RT) to generate cDNAs. During RT, the undigested amino acid left on bound RNAs leads to mutations in cDNAs or stops the RT reaction, resulting in binding site-specific marks that can be pinpointed in the downstream bioinformatics analysis. The irreversible nature of UV crosslinking enables stringent buffer conditions during IP, thus resulting in a more specific selection of targets than those captured by RIP. In addition, the use of UV light instead of formaldehyde for crosslinking captures interactions within a smaller spatial range compared to ChIP or RIP (71, 72), thus resulting in a better resolution for identification of binding targets and precise sites (Table 1).

After the first report of CLIP-seq method a decade ago (73), many modified protocols have evolved rapidly in the past few years, some of which have been utilized to characterize DEAD-box protein crosslinking sites in RNAs at transcriptome-wide scale (42, 46, 54, 66, 74–77). This

leads to a deeper and more precise understanding in the cellular functions of these RNA helicases than gene-specific studies and RIP-seq. An example that shows the more abundant information provided by CLIP-seq compared to RIP-seq is the study of RocA in enhancing RNA binding affinity of eIF4A mentioned above (66). In addition to a list of target RNAs, the sequence selectivity of eIF4A binding after RocA treatment and the specific regions in mRNAs targeted can be precisely identified using CLIP-seq but not RIP-seq (66). This also contributes to further understanding of the biochemistry mechanism of how RocA alters eIF4A activity.

CLIP-seq has been utilized to study DEAD-box proteins involved in different steps of RNA metabolism, including splicing, translation, and miRNA regulation (42, 46, 54, 74–77). One of the first studies characterizing transcriptome-wide targets of DEAD-box helicase is CLIP-seq of eIF4AIII (74), the core component of human EJC. EJC has great impact on steps in mRNA processing, including transport, translation, and decay (78). The identification of eIF4A binding sites revealed specific positions and sequence features targeted by EJC, and explained differential loading of EJC on different transcripts and exons (74). In another study using CLIP-seq for DEAD-box protein characterization, binding regions and sequence preference of DDX17 were identified in miRNAs and mRNAs from fly cells or viral RNAs, which helps to uncover the mechanism of how DDX17 restricts viral infection (42). We also used CLIP-seq for characterization of transcripts targeted by a DEAD-box protein, Dbp2, in *S. cerevisiae*, and identified mRNAs and snoRNAs as the major targets. Moreover, the binding pattern of Dbp2 along mRNAs suggests a role in transcription termination, splicing, and RNAPII pausing (see Chapter 2 and 3).

CLIP-seq of wild-type and mutant DEAD-box proteins also provides rich information about the molecular functions of these enzymes. Mutations in DEAD-box proteins could lead to loss of ATPase or RNA binding activity, or both. Depending on the activity impaired, functions of a DEAD-box helicase can be altered differentially. Particularly, a frequently found mutation of DDX3 that impairs its ATPase activity in medulloblastoma was shown to shift the binding regions as compared to the wild-type DDX3 helicase under stress conditions (75). This shift correlated with the resistance of cells to stress (75), which provides insights into the role of DDX3 mutation in tumorigenesis. CLIP-seq enables studies of changes in DEAD-box protein binding under varied conditions that would otherwise be difficult to detect in RIP-seq or ChIP-seq because of the resolution (Table 1).

By integrating the CLIP-seq data of multiple proteins, the co-regulation by DEAD-box proteins and other RBPs can also be characterized. Particularly, with enhanced CLIP (eCLIP) data of a heterogeneous nuclear ribonucleoprotein hnRNPA1 and a DEAD-box helicase DDX5, Lee et al. identified their common binding regions and a subset of mRNAs co-regulated by these two factors through alternative splicing (76). Furthermore, CLIP-seq data has been combined with other genome-wide datasets to pinpoint the specific processing step that DEAD-box proteins are involved in. For example, in addition to ChIP-seq of DDX21 that uncovered the role in transcription, Calo et al. performed CLIP-seq of DDX21 which identified the specific rRNA and snoRNA regions bound (54). These two datasets together revealed that DD21 coordinates transcriptional and post-transcriptional steps in ribosome biogenesis.

In a very recent study, CLIP-seq was also combined with transcriptome-wide probing of RNA structures inside cells (see section 1.4) and ribosome profiling to characterize the biological mechanism of how Ded1 regulates translation initiation (46). Similarly, in my research, we combined CLIP-seq and RNA structural mapping technique to study the molecular role of Dbp2 in co-transcriptional processes. This allows us to identify the direct RNA targets of Dbp2 enzymatic activity (see Chapter 2). In addition, coupled with RNA-seq analysis, we were able to correlate the binding regions with specific mRNA processing steps (see Chapter 2 & 3).

Overall, CLIP-seq not only provides valuable information about precise RNA regions targeted by DEAD-box helicase, but also complements other methods for comprehensive understanding of the molecular mechanism. However, cautions need to be taken when interpreting the data considering the limitations of the method. It has been reported that, in different protocols of CLIP-seq, nucleotide bias can arise during crosslinking by UV light (79). Moreover, since DEAD-box helicases are not processive, the transient interaction with their enzymatic targets might not be fully captured. Thus, to study the structural rearrangement of RNAs or ribonucleoprotein (RNP) complexes by DEAD-box proteins, other methods are required besides the assays for binding.

1.4 Transcriptome-wide identification of enzymatic targets of RNA helicases

Methods for mapping intracellular RNA structures can mostly be categorized into two groups: chemical modification by dimethyl sulfate (DMS) and selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) (80). DMS is a membrane permeable chemical that

methyates accessible adenines and cytosines in an RNA molecule. By probing the reactivity of these bases towards DMS, the level of base pairing can be inferred at single-nucleotide resolution (81, 82). In SHAPE-based methods, accessible ribonucleotides are acylated by SHAPE reagents at the 2' hydroxyl position in the ribose, which usually happens in flexible, single stranded regions of RNAs (83). Thus, the level of modification at each nucleotide reflects the state of RNA folding and structures, and can be measured for all four species of ribonucleotides.

Not until five years ago was the first genome-wide RNA structure probing method inside cells reported (15, 28, 84). Since then, several studies have developed methods to map secondary structures in various RNA species and regions in cells from different organisms, including yeast, plant, and human (85–89). Most of these methods combined DMS probing or SHAPE with deep sequencing. The modifications produced by DMS or SHAPE reagents can lead to reverse transcription stops or mutations in the cDNAs during library preparation. Consequently, sites of modification can be identified by the location of adaptor ligation or nucleotide conversion in sequences (Figure 1). Other methods probing long-range RNA interactions, by enriching crosslinked RNA duplexes, have also been reported (90). This advance in technology provides a new powerful tool to study the RNA/RNP structural dynamics regulated by DEAD-box proteins.

To date, only a few genome-wide studies investigated the impact of DEAD-box helicases on cellular RNA structures. One such example is the very recent study of the role of Ded1, a DEAD-box protein required for translation initiation, in RNA structure remodeling and translation efficiency in *S. cerevisiae* (46). By comparing mRNA structure profiles of the wild type and *ded1* mutant, it is found that loss of Ded1 leads to the most striking change of RNA accessibility in the 5' UTR as compared to other regions, and that the regions of structural changes align with the sites of Ded1 binding identified by CLIP-seq (46). These results collectively indicate that Ded1 facilitates translation initiation through unwinding of mRNA structures in 5' UTRs. This paradigm illustrates that combination of structural probing with other genome-wide methods is able to provide rich information on the enzymatic targets and molecular mechanisms of DEAD-box helicases inside cells. In my research, we used DMS to map RNA structures in wild type and *dbp2Δ* to pinpoint *DBP2*-dependent changes. The regions of *DBP2*-dependent structural alteration were correlated with Dbp2-binding sites identified in CLIP-seq and aligned well with annotated polyadenylation sites. By combining CLIP-seq, structural mapping, RNA-seq, and RNAPII ChIP-seq, we provide molecular mechanism of how Dbp2 may function in transcriptional termination

(see Chapter 2), the first example illustrating DEAD-box proteins remodels transcriptome-wide RNA/RNP structures to regulate nuclear RNA metabolism.

Nevertheless, challenges and caveats remain in probing RNA/RNP structural dynamics that is dependent on RNA helicases. First, methods for nascent RNA structural probing during transcription have not been established in eukaryotes although it has been reported in bacteria (91). Structures in nascent RNAs, such as introns, can be determinants of splicing efficiency and exon inclusion (3–9), impacting the downstream steps in gene expression including export and 3' processing (92, 93). As the enzymatic activities of DEAD-box helicases may be essential to co-transcriptional processes (41, 94, 95), a technique analyzing nascent RNAs can provide invaluable insight into DEAD-box protein functions in co-transcriptional processes. Second, most structural probing techniques only provide information of local secondary structures. Although efforts have been devoted to mapping of long-range RNA duplex formation (90), tertiary RNA structures in living cells remain uncharacterized. Previously, DEAD-box helicases have been shown to assist formation of functional tertiary structures through remodeling secondary structures (96). Therefore, experimental and computational techniques predicting tertiary RNA folding will help to understand the role of DEAD-box helicase in formation of complex RNA and RNP structures. Finally, caution should be taken when interpreting the changes derived from chemical probing experiments. Regions inaccessible to chemical modification can be protected by RNA secondary and tertiary structures or protein binding. Therefore, signal changes dependent on DEAD-box helicases can originate from direct RNA/RNP structure remodeling or indirect impacts on the RNA-protein complex compositions (Figure 1).

1.5 Concluding remarks

Genome-wide approaches have enabled broad identification of intracellular targets for DEAD-box proteins. Chemical probing of RNA structures combined with deep sequencing also opens a new avenue for studies of enzymatic targets for these helicases at transcriptome scale. Although our understanding in the biological functions of DEAD-box proteins is greatly advanced with these high-throughput tools, many gaps in the field remain to be filled.

The relationship between detected RNA binding and the direct biological role of certain DEAD-box proteins is still largely uncharacterized. Transcripts encoding proteins in specific biological processes or pathways are often enriched in DEAD-box protein binding RNAs (74, 77,

97). However, it is rarely discussed how this binding specificity is conferred and how this binding preference may regulate specific cellular pathways. It would be interesting to investigate whether certain properties, including sequences and structures, in transcripts of the targeted processes contribute to the specific targeting by DEAD-box proteins. In addition, whether the loss of binding results in change in the expression of factors in certain pathways is important to further understand the cellular function of DEAD-box proteins.

Since DEAD-box proteins display a variety of biochemical activities, it would be critical to be able to uncouple the requirement of different activities for their specific cellular functions. For example, DDX5 has been shown to bind a lncRNA, Hox transcript antisense intergenic RNA (HOTAIR), and stabilize polycomb repressive complex 2 (PRC2). Loss of RNA unwinding activity does not affect the stabilization of a component in PRC2, SUZ12, by DDX5 (D248N) (98). On the other hand, the helicase activity of DDX5 is required for the inclusion of *tau* pre-mRNA exon 10 (41). This example illustrates that diverse biochemical activities of DEAD-box proteins could have differential impacts in different processes. Thus, comparison and integration of high-throughput datasets, such as binding and structural remodeling data, can be invaluable for dissecting molecular roles of DEAD-box proteins in different steps of gene expression (Figure 1.2).

1.6 References

1. Mortimer SA, Kidwell MA, Doudna JA. 2014. Insights into RNA structure and function from genome-wide studies. *Nat Rev Genet*.
2. Lewis CJT, Pan T, Kalsotra A. 2017. RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat Rev Mol Cell Biol*.
3. Mueller N, van Bel N, Berkhout B, Das AT. 2014. HIV-1 splicing at the major splice donor site is restricted by RNA structure. *Virology* 468:609–620.
4. Singh NN, Singh RN, Androphy EJ. 2007. Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res* 35:371–389.
5. Blanchette M, Chabot B. 1997. A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *RNA* 3:405–419.

6. Watakabe A, Inoue K, Sakamoto H, Shimura Y. 1989. A secondary structure at the 3' splice site affects the in vitro splicing reaction of mouse immunoglobulin μ chain pre-mRNAs. *Nucleic Acids Res* 17:8159–8169.
7. Charpentier B, Rosbash M. 1996. Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. *RNA* 2:509–22.
8. Muh SJ, Hovhannisyan RH, Carstens RP. 2002. A non-sequence-specific double-stranded RNA structural element regulates splicing of two mutually exclusive exons of fibroblast growth factor receptor 2 (FGFR2). *J Biol Chem* 277:50143–50154.
9. Kreahling JM, Graveley BR. 2005. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila* Dscam pre-mRNA. *Mol Cell Biol* 25:10251–10260.
10. Pelletier J, Sonenberg N. 1985. Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* 40:515–526.
11. Pelletier J, Sonenberg N. 1987. The involvement of mRNA secondary structure in protein synthesis. *Biochem Cell Biol* 65:576–581.
12. Tuller T, Waldman YY, Kupiec M, Ruppin E. 2010. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci* 107:3645–3650.
13. Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, Mavrakis KJ, Jiang M, Roderick JE, Van der Meulen J, Schatz JH, Rodrigo CM, Zhao C, Rondou P, de Stanchina E, Teruya-Feldstein J, Kelliher MA, Speleman F, Porco JA, Pelletier J, R  tsch G, Wendel HG. 2014. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* 513:65–70.
14. Incarnato D, Neri F, Anselmi F, Oliviero S. 2014. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol* 15:491.
15. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505:696–700.

16. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, Chang HY. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505:706–709.
17. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467:103–107.
18. Rodnina M V. 2016. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci.*
19. Faure G, Ogurtsov AY, Shabalina SA, Koonin E V. 2017. Adaptation of mRNA structure to control protein folding. *RNA Biol.*
20. Faure G, Ogurtsov AY, Shabalina SA, Koonin E V. 2016. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res* 44:10898–10911.
21. Geisberg J V., Moqtaderi Z, Fan X, Oszolak F, Struhl K. 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* 156:812–824.
22. Li X, Quon G, Lipshitz HD, Morris Q. 2010. Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA* 16:1096–107.
23. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. Understanding the transcriptome through RNA structure. *Nat Rev Genet* 12 VN-r:641–655.
24. Dienstbier M, Boehl F, Li X, Bullock SL. 2009. Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. *Genes Dev* 23:1546–1558.
25. Bullock SL, Ringel I, Ish-Horowicz D, Lukavsky PJ. 2010. A'-form RNA helices are required for cytoplasmic mRNA transport in *Drosophila*. *Nat Struct Mol Biol* 17:703–709.
26. Schroeder R, Barta A, Semrad K. 2004. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* 5:908–919.
27. Doetsch M, Schroeder R, Fürtig B. 2011. Transient RNA-protein interactions in RNA folding. *FEBS J.*
28. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505:701–5.
29. Bourgeois CF, Mortreux F, Auboeuf D. 2016. The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat Rev Mol Cell Biol* 17:426–438.
30. Jarmoskaite I, Russell R. 2014. RNA Helicase Proteins as Chaperones and Remodelers. *Annu Rev Biochem* 83:697–725.

31. Fairman-Williams ME, Guenther UP, Jankowsky E. 2010. SF1 and SF2 helicases: Family matters. *Curr Opin Struct Biol*.
32. Russell R, Jarmoskaite I, Lambowitz AM. 2013. Toward a molecular understanding of RNA remodeling by DEAD-box proteins. *RNA Biol* 10:44–55.
33. Rudolph MG, Klostermeier D. 2015. When core competence is not enough: Functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. *Biol Chem*.
34. Fuller-Pace F V. 2013. DEAD box RNA helicase functions in cancer. *RNA Biol* 10:121–132.
35. Steimer L, Klostermeier D. 2012. RNA helicases in infection and disease. *RNA Biol*.
36. Putnam AA, Jankowsky E. 2013. DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim Biophys Acta - Gene Regul Mech*.
37. Liebeg A, Mayer O, Waldsich C. 2010. DEAD-box protein facilitated RNA folding in vivo. *Rna Biol* 7:803–811.
38. Ruminski DJ, Watson PY, Mahen EM, Fedor MJ. 2016. A DEAD-box RNA helicase promotes thermodynamic equilibration of kinetically trapped RNA structures in vivo. *RNA* 22:416–427.
39. Potratz JP, Del Campo M, Wolf RZ, Lambowitz AM, Russell R. 2011. ATP-dependent roles of the DEAD-box protein Mss116p in group II intron splicing in vitro and in vivo. *J Mol Biol* 411:661–679.
40. Dardenne E, PolayEspinoza M, Fattet L, Germann S, Lambert MP, Neil H, Zonta E, Mortada H, Gratadou L, Deygas M, Chakrama F, Samaan S, Desmet FO, Tranchevent LC, Dutertre M, Rimokh R, Bourgeois CF, Auboeuf D. 2014. RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Rep* 7:1900–1913.
41. Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, Liu Z, Chen S, Wu JY. 2011. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol* 31:1812–1821.
42. Moy RH, Cole BS, Yasunaga A, Gold B, Shankarling G, Varble A, Molleston JM, Tenover BR, Lynch KW, Cherry S. 2014. Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell* 158:764–777.

43. Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, Liu Z, Chen S, Wu JY. 2011. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol* 31:1812–1821.
44. Svitkin Y V., Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham GJ, Sonenberg N. 2001. The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* 7:382–394.
45. Soto-Rifo R, Rubilar PS, Limousin T, De Breyne S, Décimo D, Ohlmann T. 2012. DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J* 31:3745–3756.
46. Guenther UP, Weinberg DE, Zubradt MM, Tedeschi FA, Stawicki BN, Zagore LL, Brar GA, Licatalosi DD, Bartel DP, Weissman JS, Jankowsky E. 2018. The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature* 559:130–134.
47. Linder P, Jankowsky E. 2011. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12:505–516.
48. Oberer M, Marintchev A, Wagner G. 2005. Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev* 19:2212–2223.
49. Young CL, Khoshnevis S, Karbstein K. 2013. Cofactor-dependent specificity of a DEAD-box protein. *Proc Natl Acad Sci U S A* 110:E2668-76.
50. Schutz P, Bumann M, Oberholzer AE, Bieniossek C, Trachsel H, Altmann M, Baumann U. 2008. Crystal structure of the yeast eIF4A-eIF4G complex: An RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci* 105:9564–9569.
51. Le Hir H, Andersen GR. 2008. Structural insights into the exon junction complex. *Curr Opin Struct Biol*.
52. Fairman ME, Maroney PA, Wang W, Bowers HA, Gollnick P, Nilsen TW, Jankowsky E. 2004. Protein displacement by DExH/D “RNA helicases” without duplex unwinding. *Science* 304:730–4.
53. Tran EJ, Zhou Y, Corbett AH, Wentz SR. 2007. The DEAD-Box Protein Dbp5 Controls mRNA Export by Triggering Specific RNA:Protein Remodeling Events. *Mol Cell* 28:850–859.
54. Calo E, Flynn RA, Martin L, Spitale RC, Chang HY, Wysocka J. 2014. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* 518:249–253.

55. Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, Felsenfeld G. 2010. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev* 24:2543–2555.
56. Song C, Hotz-Wagenblatt A, Voit R, Grummt I. 2017. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev* 31:1370–1381.
57. Fuller-Pace F V. 2006. DExD/H box RNA helicases: Multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.*
58. Sarkar M, Ghosh MK. 2016. DEAD box RNA helicases: crucial regulators of gene expression and oncogenesis. *Front Biosci* 21:4386.
59. Jensen ED, Niu L, Caretti G, Nicol SM, Teplyuk N, Stein GS, Sartorelli V, Van Wijnen AJ, Fuller-Pace F V., Westendorf JJ. 2008. p68 (Ddx5) interacts with Runx2 and regulates osteoblast differentiation. *J Cell Biochem* 103:1438–1451.
60. Skourti-Stathaki K, Proudfoot NJ. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.*
61. Clark EL, Hadjimichael C, Temperley R, Barnard A, Fuller-Pace F V., Robson CN. 2013. p68/Ddx5 Supports β -Catenin & RNAP II during Androgen Receptor Mediated Transcription in Prostate Cancer. *PLoS One* 8.
62. Zhang L, Yang Y, Li B, Scott IC, Lou X. 2018. The DEAD-box RNA helicase Ddx39ab is essential for myocyte and lens development in zebrafish. *Development* 145:dev161018.
63. Das M, Renganathan A, Dighe SN, Bhaduri U, Shettar A, Mukherjee G, Kondaiah P, Satyanarayana Rao MR. 2018. DDX5/p68 associated lncRNA LOC284454 is differentially expressed in human cancers and modulates gene expression. *RNA Biol* 15:214–230.
64. Zhou F, Liu Y, Rohde C, Pauli C, Gerloff D, Köhn M, Misiak D, Bäumer N, Cui C, Göllner S, Oellerich T, Serve H, Garcia-Cuellar M-P, Slany R, Maciejewski JP, Przychodzen B, Seliger B, Klein H-U, Bartenhagen C, Berdel WE, Dugas M, Taketo MM, Farouq D, Schwartz S, Regev A, Hébert J, Sauvageau G, Pabst C, Hüttelmaier S, Müller-Tidow C. 2017. AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. *Nat Cell Biol* 19:844–855.

65. Miller JE, Zhang L, Jiang H, Li Y, Pugh BF, Reese JC. 2017. Genome-Wide Mapping of Decay Factor-mRNA Interactions in Yeast Identifies Nutrient Responsive Transcripts as Targets of the Deadenylase Ccr4. *G3* g3.300415.2017.
66. Iwasaki S, Floor SN, Ingolia NT. 2016. Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature* 534:558–561.
67. Andreou AZ, Klostermeier D. 2013. The DEAD-box helicase eIF4A. *RNA Biol* 10:19–32.
68. Arun G, Akhade VS, Donakonda S, Rao MRS. 2012. mrhl RNA, a Long Noncoding RNA, Negatively Regulates Wnt Signaling through Its Protein Partner Ddx5/p68 in Mouse Spermatogonial Cells. *Mol Cell Biol* 32:3140–3152.
69. Jung C, Mittler G, Oswald F, Borggrefe T. 2013. RNA helicase Ddx5 and the noncoding RNA SRA act as coactivators in the Notch signaling pathway. *Biochim Biophys Acta - Mol Cell Res* 1833:1180–1189.
70. Andersen CBF, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CLP, Pedersen JS, Séraphin B, Hir H Le, Andersen GR. 2006. Structure of the exon junction core complex with a trapped DEAD-Box ATPase bound to RNA. *Science* (80-) 313:1968–1972.
71. Pashev IG, Dimitrov SI, Angelov D. 1991. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem Sci* 16:323–326.
72. Hoffman EA, Frey BL, Smith LM, Auble DT. 2015. Formaldehyde crosslinking: A tool for the study of chromatin complexes. *J Biol Chem*.
73. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456:464–469.
74. Saulière J, Murigneux V, Wang Z, Marquet E, Barbosa I, Le Tonquèze O, Audic Y, Paillard L, Crollius HR, Le Hir H. 2012. CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat Struct Mol Biol* 19:1124–1131.
75. Oh S, Flynn RA, Floor SN, Purzner J, Martin L, Do BT, Schubert S, Vaka D, Morrissy S, Li Y, Kool M, Hovestadt V, Jones DTW, Northcott PA, Risch T, Warnatz H-J, Yaspo M-L, Adams CM, Leib RD, Breese M, Marra MA, Malkin D, Lichter P, Doudna JA, Pfister SM, Taylor MD, Chang HY, Cho Y-J. 2016. Medulloblastoma-associated DDX3 variant selectively alters the translational response to stress. *Oncotarget* 7:28169–28182.

76. Lee YJ, Wang Q, Rio DC. 2018. Coordinate regulation of alternative pre-mRNA splicing events by the human RNA chaperone proteins hnRNPA1 and DDX5. *Genes Dev* 32:1060–1074.
77. Zhong W, Li Z, Zhou M, Xu T, Wang Y. 2018. DDX1 regulates alternative splicing and insulin secretion in pancreatic β cells. *Biochem Biophys Res Commun* 500:751–757.
78. Hir H Le, Saulière J, Wang Z. 2016. The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Biol*.
79. Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M, Ule J. 2012. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* 13:R67.
80. Bevilacqua PC, Ritchey LE, Su Z, Assmann SM. 2016. Genome-Wide Analysis of RNA Secondary Structure. *Annu Rev Genet* 50:235–266.
81. Inoue T, Cech TR. 1985. Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci* 82:648–652.
82. Lempereur L, Nicoloso M, Riehl N, Ehresmann C, Ehresmann B, Bachellerie JP. 1985. Conformation of yeast 18S rRNA. Direct chemical probing of the 5' domain in ribosomal subunits and in deproteinized RNA by reverse transcriptase mapping of dimethyl sulfate-accessible sites. *Nucleic Acids Res* 13:8339–8357.
83. Wilkinson KA, Merino EJ, Weeks KM. 2006. Quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 1:1610–1616.
84. Wan Y, Qu K, Zhang QC, Flynn R a, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, Chang HY. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505:706–9.
85. Talkish J, May G, Lin Y, Woolford JL, McManus CJ. 2014. Mod-seq: High-throughput sequencing for chemical probing of RNA structure. *RNA* 20:713–720.
86. Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM. 2015. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat Protoc* 10:1050–1066.

87. Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2016. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* 14:75–82.
88. Ritchey LE, Su Z, Tang Y, Tack DC, Assmann SM, Bevilacqua PC. 2017. Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. *Nucleic Acids Res* 45:e135.
89. Smola MJ, Weeks KM. 2018. In-cell RNA structure probing with SHAPE-MaP. *Nat Protoc* 13:1181–1195.
90. Gong J, Ju Y, Shao D, Zhang QC. 2018. Advances and challenges towards the study of RNA-RNA interactions in a transcriptome-wide scale. *Quant Biol*.
91. Incarnato D, Morandi E, Anselmi F, Simon LM, Basile G, Oliviero S. 2017. In vivo probing of nascent RNA structures reveals principles of cotranscriptional folding. *Nucleic Acids Res* 45:9716–9725.
92. Reed R. 2003. Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol*.
93. Kaida D. 2016. The reciprocal regulation between splicing and 3'-end processing. *Wiley Interdiscip Rev RNA*.
94. Xing Z, Ma WK, Tran EJ. 2018. The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscip Rev RNA* 0:e1519.
95. Zonta E, Bittencourt D, Samaan S, Germann S, Dutertre M, Auboeuf D. 2013. The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. *Nucleic Acids Res* 41:554–564.
96. Pan C, Potratz JP, Cannon B, Simpson ZB, Ziehr JL, Tijerina P, Russell R. 2014. DEAD-Box Helicase Proteins Disrupt RNA Tertiary Structure Through Helix Capture. *PLoS Biol* 12.
97. Valentin-Vega YA, Wang YD, Parker M, Patmore DM, Kanagaraj A, Moore J, Rusch M, Finkelstein D, Ellison DW, Gilbertson RJ, Zhang J, Kim HJ, Taylor JP. 2016. Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. *Sci Rep* 6.

98. Zhang H, Xing Z, Mani SKK, Bancel B, Durantel D, Zoulim F, Tran EJ, Merle P, Andrisani O. 2016. RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis. *Hepatology* 64:1033–1048.

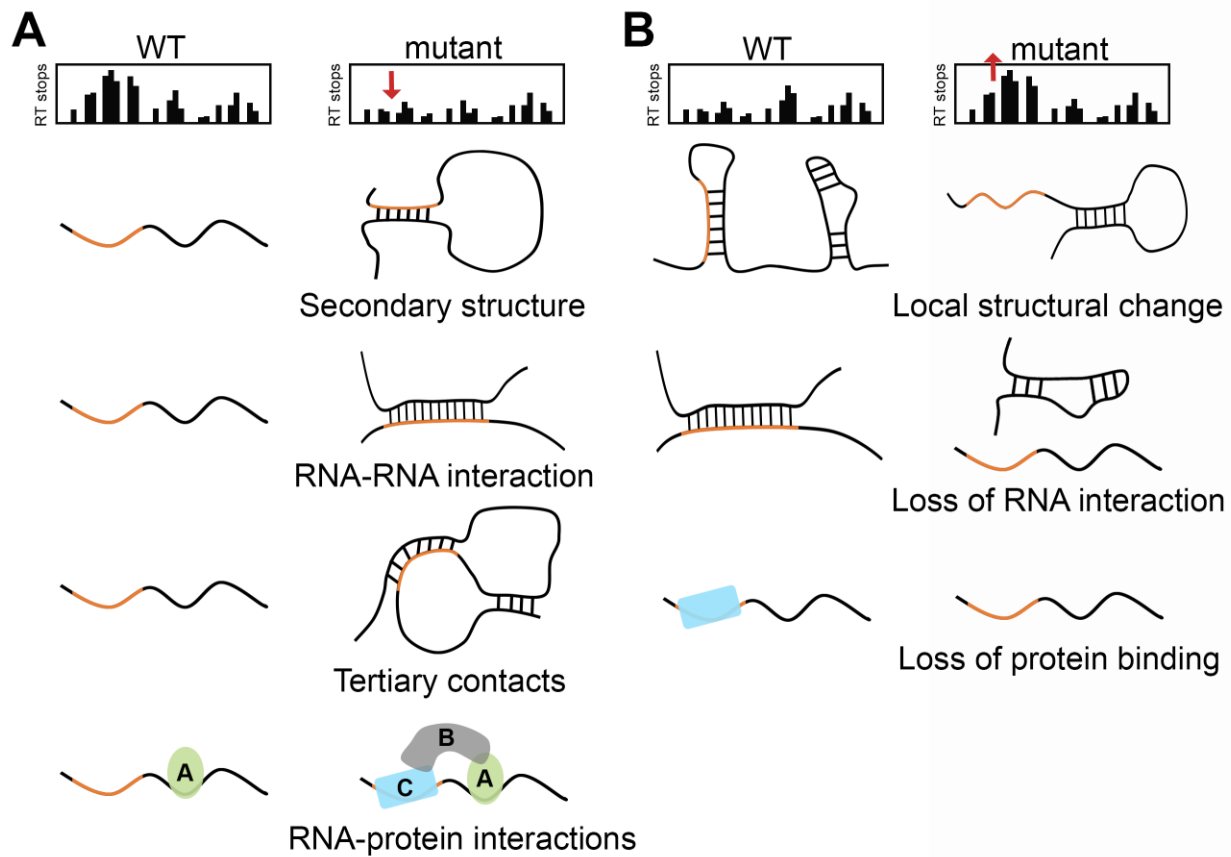


Figure 1.1. Potential scenarios for (A) decrease and (B) increase in RNA structural probing signals in DEAD-box protein mutant strains as compared to wild-type cells.

Both directions of changes can be observed in the same transcript. The orange region in the schematic representation corresponds to the regions with detected signal changes. (A) The decrease in the RT stops indicates reduced accessibility to chemical modifications in the mutant versus wild type. This can be a result of formation of secondary structures, inter-molecule interaction between two transcripts, folding of tertiary structures, and binding by RBPs in the absence of active DEAD-box helicases. (B) The increase in the RT stops indicates higher accessibility to chemical modifications in the mutant versus wild type. This can be contributed by the change in local structural changes, loss of RNA-RNA interaction, and reduced protein binding in cells lacking functional DEAD-box helicases.

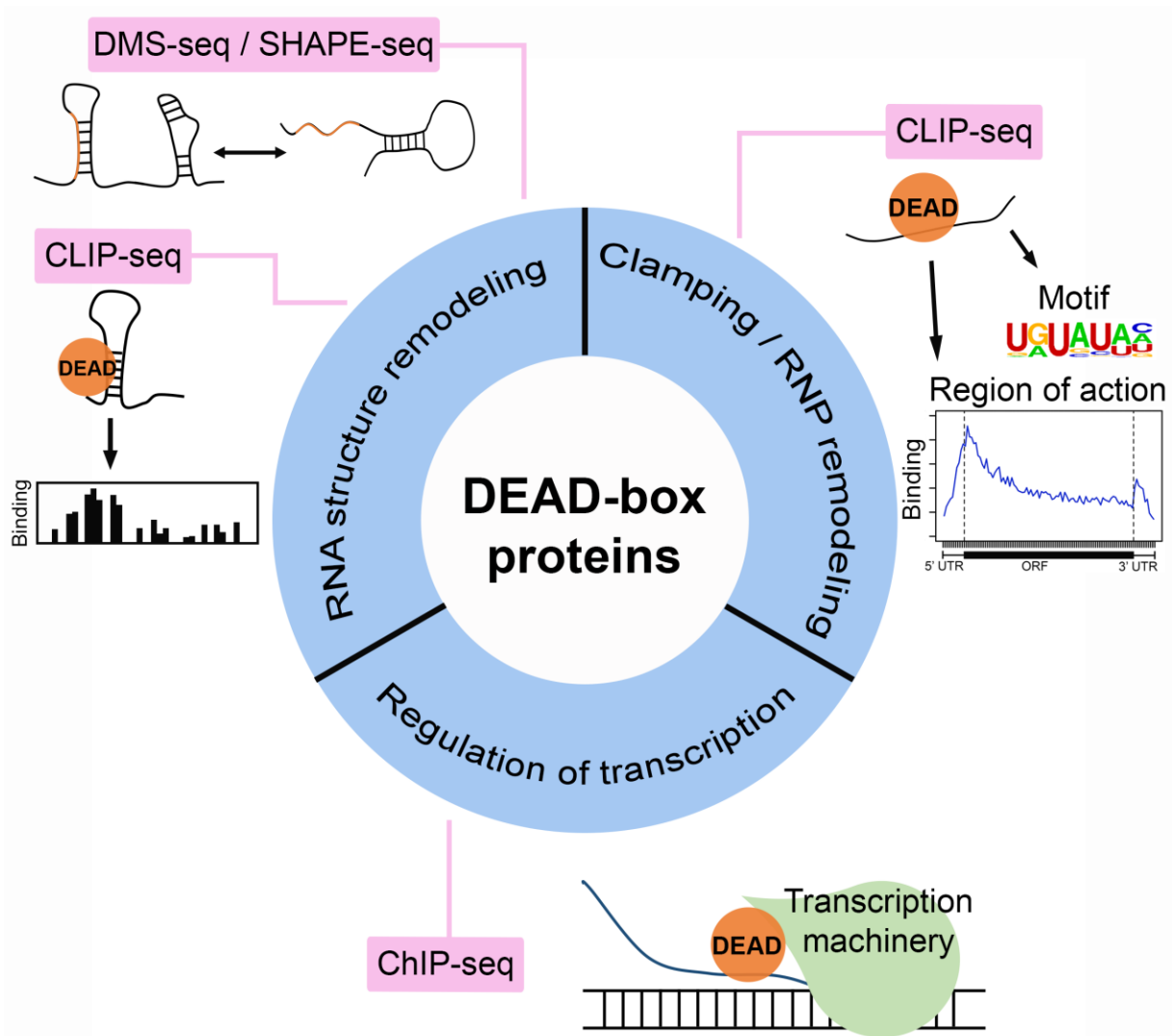


Figure 1.2. Genome-wide techniques for studying cellular functions of DEAD-box proteins.

Chemical probing of transcriptome-wide RNA structural changes dependent on specific DEAD-box helicase can be integrated with the binding data derived from CLIP-seq to determine the precise target sites of structural remodeling. CLIP-seq can also be used to probe the sites of clamping or RNPs that are controlled by the target DEAD-box protein. Additionally, using ChIP-seq, the distribution of DEAD-box proteins across the genome can be mapped to further the understanding of their co-transcriptional functions.

Table 1.1. Genome-wide methods for detection of DEAD-box protein binding on nucleic acids

	ChIP-seq	RIP-seq	CLIP-seq
Crosslinking method	Formaldehyde	None or formaldehyde	UV light
Contact range captured	~ 2Å	~ 2Å	Direct contact
Nucleic acids purified	DNA	RNA	RNA
Resolution	~ 200 bp	20 – 70 nt	Single nucleotide
Information gained	Genomic loci targeted by DEAD-box proteins	RNA transcripts interacting with DEAD-box proteins	Specific sites in RNAs bound by DEAD-box proteins
Major disadvantage	Signal could come from various sources, including the interaction between DEAD-box protein and DNA, nascent RNA, or transcription machinery	Precise binding sites of DEAD-box proteins (6-8 nt footprint) cannot be determined	UV crosslinking could result in sequence bias
References	(54, 55)	(54, 62, 63, 65, 66)	(14, 18, 54, 66, 74–77)

CHAPTER 2. GENOME-WIDE DISCOVERY OF DEAD-BOX RNA HELICASE TARGETS REVEALS RNA STRUCTURAL REMODELING IN TRANSCRIPTION TERMINATION

2.1 Introduction

RNA helicases are found in all kingdoms of life, playing central roles in all aspects of RNA metabolism (Bourgeois et al., 2016). Among them, DEAD-box proteins constitute the largest RNA helicase family. These enzymes have a conserved helicase core which is responsible for ATP binding, hydrolysis, and RNA binding, and are characterized by the Asp-Glu-Ala-Asp (D-E-A-D) motif. Most DEAD-box helicases are involved in steps of gene expression in cells, including transcription to translation to decay (Linder and Jankowsky, 2011). However, the detailed molecular actions of these helicases remain to be characterized.

Most DEAD-box helicases have an ATP-dependent RNA-unwinding activity *in vitro* (Putnam and Jankowsky, 2013). This activity catalyzes a wide variety of biochemically distinct actions including non-processive, RNA duplex unwinding (Rogers et al., 1999; Yang and Jankowsky, 2006), RNA-protein complex (RNP) remodeling activity *in vitro* (Fairman et al., 2004; Tran et al., 2007), and ATP-dependent “clamping” of multiprotein complexes onto RNA (Ballut et al., 2005; Nielsen et al., 2008). Studies have also shown that DEAD-box helicases can act as chaperones to promote RNA folding both *in vitro* and *in vivo* (Liebeg et al., 2010; Potratz et al., 2011; Tijerina et al., 2006; Yang and Jankowsky, 2005). For example, Mss116 in *Saccharomyces cerevisiae* assists the folding of functional group I and II introns by unwinding misfolded RNAs to allow exchange between kinetically trapped, nonfunctional structures and functional conformations (Liebeg et al., 2010; Potratz et al., 2011). Human DEAD-box helicases including DDX5 and DDX17 have also been reported to unwind secondary RNA structures and thereby regulate alternative splicing (Dardenne et al., 2014; Kar et al., 2011). RNA remodeling activity also appears to be critical for translation, as a recent genome-wide study of the translation factor and DEAD-box helicase Ded1 showed that this enzyme resolves structures in the 5' ends of genes and controls translational start site choice in *S. cerevisiae* (Guenther et al., 2018). These examples implicate DEAD-box helicases as potential regulators of RNA metabolism and gene expression, yet we lack a thorough understanding of how RNA structure and RNA-protein complex assembly impacts basic molecular steps within these processes.

For proper gene expression, the basic steps in transcription include initiation, elongation, and termination. Termination by RNA polymerase II (RNAPII) is mediated mainly by two complexes in *S. cerevisiae*: the cleavage and polyadenylation (CPF) complex and the Nrd1-Nab3-Sen1 (NNS) complex. CPF-dependent 3' end processing is the primary mode of termination for messenger RNAs, whereas the NNS complex, a trimeric assembly of RNA-binding proteins Nrd1 and Nab3 with the RNA-DNA helicase Sen1, promotes termination of short, non-coding RNAPII transcripts and some mRNAs (Porrua and Libri, 2015; Rondón et al., 2009). The NNS complex has also been implicated in “failsafe” termination, whereby NNS target sites can rescue defective termination from an upstream CPF-dependent site to prevent aberrant gene expression (Rondón et al., 2009).

Previous results from our lab showed that the ortholog of DDX5 in *S. cerevisiae*, Dbp2 (Xing et al., 2018), is required for efficient termination of RNAPII transcription, as loss of *DBP2* results in accumulation of a 3' extended *GAL10* mRNA and *GAL10* long non-coding (lnc) RNA (Cloutier et al., 2012). Both Dbp2 and DDX5 exhibit highly efficient RNA duplex unwinding *in vitro*, consistent with a role in altering secondary structure (Ma et al., 2013; Xing et al., 2017). Furthermore, Dbp2 associates with actively transcribed chromatin in an RNA-dependent manner (Ma et al., 2016) and is required for pre-mRNA maturation and mRNP assembly, as evidenced by reduced binding of export factors Nab2, Yra1 and Mex67 in *dbp2Δ* cells (Ma et al., 2013). As efficient termination is necessary for proper assembly of mRNA export factors (Qu et al., 2009), these two steps are likely linked through a common, upstream biochemical step mediated by Dbp2.

Several recent studies combining classic chemical mapping techniques with next-generation sequencing have shown that mRNAs are largely less structured *in vivo* than *in vitro*, that secondary structure alteration by single nucleotide polymorphisms may underlie human diseases, and that these changes in structure, when overlapping regulatory sites, have the potential to provide a new level of gene regulation (Corley et al., 2015; Ding et al., 2014; Rouskin et al., 2013; Wan et al., 2014). In addition, early biochemical studies have shown that the formation of RNA secondary structures in polyadenylation signal elements inhibits 3' processing and termination factor binding of pre-mRNAs in mammalian cells (Chen and Wilusz, 1998; Klasens et al., 1998). Genome-wide RNA structure mapping in *Arabidopsis* also revealed widespread formation of secondary structures in 3' UTRs, suggestive of a broad role for secondary structure in termination (Ding et al., 2014). Nevertheless, a role for cellular RNA structure remodeling in transcriptional termination

has not been established to date. Using a combination of RNA sequencing-based techniques combined with classic yeast genetics and molecular biology, we provide evidence that the DEAD-box RNA helicase Dbp2 in *S. cerevisiae* remodels secondary structures within the 3' ends of a subset of mRNAs to promote efficient termination by the NNS complex. This reveals that DEAD-box RNA helicases remodel mRNA structure *in vivo* and that structural alteration is essential for proper gene expression.

2.2 Materials and Methods

Yeast strains and plasmids

The strains and plasmids used in this study were constructed using classical yeast genetic and/or cloning methods and are listed in Table 2.1 and 2.2. To construct the termination reporter plasmids, a DNA fragment containing the *YOP1* 3' UTR sequence (chrXVI, 624203 - 624342) was generated from genomic DNA and inserted into the XhoI site of pGAC24 to make pGAC24-YOP1. The *YOP1* 3'UTR of the resulting plasmid was mutated by site-directed mutagenesis of the pGAC24-YOP1 construct. The primers for strain construction and cloning are listed in Table 2.3.

Procedures of iCLIP-seq library construction

a. Purification of Dbp2-bound RNAs

iCLIP-seq was adapted from the FAST-iCLIP protocol (Flynn et al., 2015) with the following modifications. *DBP2-3XFLAG* strains were grown in YP with 2% glucose at 30 °C to an OD_{600nm} of 0.5~0.7. 250 mL of cells were harvested and washed in ice-cold TBS with 2% glucose. After centrifugation, pellets were resuspended in 12 mL of ice-cold TBS with 2% glucose and irradiated on ice by UV-C (254 nm) at 180 mJ/cm² twice for 2.5 min, with a 45-second rest. Cells were subsequently harvested, frozen in liquid nitrogen, and lysed cryogenically using a Retsch Oscillating Mill MM400. Lysed cells were re-suspended in the CLIP Lysis Buffer as previously described (Flynn et al., 2015). The soluble fraction was digested with TURBO DNase (ThermoFisher Scientific) and RNase I (ThermoFisher Scientific) at 37 °C for 10 min. RNase-treated lysates were incubated with FLAG M2 (Sigma) antibody-conjugated Protein G Dynabeads (ThermoFisher) at 4 °C for 2 hours and then washed as described (Flynn et al., 2015). The labeling and purification of immunoprecipitated RNAs were performed as described (Flynn et al., 2015).

b. Preparation of cDNAs

RNAs were mixed with barcoded 5' phosphorylated RT primers (Table 2.4) and heated at 65 °C for 2 min. The cDNAs were generated using TGIRT-III enzyme according to the manufacturer's instruction. After reverse transcription (RT), RNAs were degraded with RNase A and RNase H at 37 °C for 30 min. The RT products were circularized and purified as described (Flynn et al., 2015).

c. Library amplification and sequencing

Real time PCR reactions were set up using 2X SYBR master mix (Applied Biosystems), P3 and P5_short primers. After 40 cycles of PCR, the products were separated by denaturing PAGE and stained with SYBR Gold (ThermoFisher Scientific). DNAs larger than 75 nt were extracted from the gel. After elution, DNAs were ethanol-precipitated and re-suspended in water. PCR products were further amplified by 10 cycles of PCR with P3 and P5_Solexa primers (Table 2.4). The amplified products above 140 nts were gel purified and re-suspended in water. All the samples were analyzed by Agilent Bioanalyzer to determine the size distribution of the library. Sequencing was performed on the illumina Hiseq 2500 platform for 2 x 100 bp paired-end cycle run.

d. Processing of iCLIP-seq data

Solexa adaptors were removed using Trimmomatic (version 0.36) (Bolger et al., 2014). The FASTQ files were de-multiplexed and PCR duplicates were removed based on the random barcodes incorporated in the RT primers using scripts provided by the Chang lab (<https://github.com/qczhang/icSHAPE/tree/master/scripts>, (Flynn et al., 2016)). Barcode sequences (13 nt) were trimmed from the 5' end of retained forward reads using cutadapt (version 1.9.1). The processed reads were mapped to the S288C reference genome (R64-2-1, from Saccharomyces Genome Database) using STAR (version 2.5.2b) (Dobin et al., 2013). Reads mapped to one or two sites were kept for the following analysis to include transcripts from duplicated genes, including rRNAs. The mapping rate of each replicate was > 95%. The number of reads mapped to each annotated RNA transcript were counted using the summarizeOverlaps function in the Bioconductor package "GenomicAlignments" (version 1.8.4). For each of the three Dbp2-iCLIP replicates, reads that did not overlap with any read in the other two replicates were regarded as background and were discarded. Transcripts that had less than 5 counts in each library were filtered from the analysis. Only transcripts that were identified in all three replicates were regarded as binding targets. The nucleotide position before the start of each read was extracted from the forward reads as the crosslinking site in each replicate. Raw reads of Dbp2-iCLIP-seq were deposited to Gene Expression Omnibus (GEO): GSE106479.

RNAPII chromatin immunoprecipitation and sequencing (ChIP-seq)

ChIP was performed as described (Cloutier et al., 2013) using anti-FLAG M2 monoclonal antibody (Sigma-Aldrich) for immunoprecipitation of endogenously 3XFLAG-tagged Rpb3. Sequencing libraries were prepared from the input and immunoprecipitated DNAs using NEXTflex ChIP-Seq Kit (BIOO Scientific, Austin, TX) according to the manufacturer's instructions. All the libraries were analyzed by Agilent Bioanalyzer to determine the size distribution. Sequencing was performed on the illumina Miseq platform for 2 x 150 bp paired-end cycle run.

Processing of RNAPII ChIP-seq data

Adaptor sequences were removed using Trimmomatic (v 0.36) (Bolger et al., 2014). The processed reads were mapped to the S288C reference genome (R64-2-1, from *Saccharomyces* Genome Database) using Bowtie 2 (v 2.3.3.1) (Langmead and Salzberg, 2012). Peaks of RNAPII were then determined by MACS2 (v 2.1.2) (Zhang et al., 2008). The fold enrichment of each peak over the background signal was calculated using the 'bdgcmp' function in MACS2 and presented as normalized RNAPII occupancy. The overall RNAPII occupancy around termination sites of snoRNAs or mRNAs was analyzed by deepTools (v 3.1.1) (Ramírez et al., 2016).

Identification of Dbp2 binding motifs in mRNA 3' UTRs

To identify the enriched sequence motifs in close proximity of Dbp2 binding sites in mRNA 3' UTRs, each binding site was extended by 5 nts on each side (total 11 nt in length). All the binding regions derived from the three replicates were combined into one bed file and were analyzed by HOMER 4.7b (Heinz et al., 2010) using the findMotifsGenome.pl function.

Metagene analysis of Dbp2-binding sites in mRNAs

The coordinates of the untranslated regions (UTRs) for each mRNA were derived from two published datasets (Nagalakshmi et al., 2008; Yassour et al., 2009). If the UTR coordinates for the same transcript were different in the two datasets, the coordinates that had the widest range were used. For mRNAs without UTR annotations, 135 nts (close to the median lengths of all yeast 5' and 3' UTRs) were added before and after the open reading frame (ORF) region as 5' and 3' UTRs. The metagene plots were generated using custom R scripts with the following steps. Each 5' UTR,

ORF, and 3' UTR, was divided into 10, 80, and 10 bins, respectively. The number of bins is proportional to the median length of each region in analyzed mRNAs. The number of crosslinking counts in each bin was then normalized to the library size of each replicate and to the expression level to obtain reads per kilobase of transcript per million mapped reads (RPKM units) of each transcript based on the Structure-seq data (wild type, no DMS treatment). The normalized Dbp2 occupancy in each bin was then divided by the total occupancy in all bins to calculate the distribution of Dbp2. The derived value for each bin in each transcript was averaged among the three replicates, and then averaged across all the transcripts analyzed. The distribution of Dbp2 occupancy across a set of mRNAs was plotted as a line graph based on the calculated value in each bin.

Genomic Localization of Dbp2 and Rna15 (or Nrd1/Nab3) binding sites

The RNA binding data of Rna15, Nrd1, and Nab3 were downloaded from the Gene Expression Omnibus (GEO) database (Rna15: GSM1442555, Nrd1: GSM791764, Nab3: GSM791767) (Baejen et al., 2014; Creamer et al., 2011). For each protein binding site, the distance between the position and the closest Dbp2 binding site on the same transcript was calculated. For each nearest pair of Rna15 or Nrd1 or Nab3 with Dbp2, the occupancy of the protein was normalized to the occupancy of Dbp2. The average occupancy of each protein at each distance was then calculated and plotted to demonstrate the pattern of the protein distribution near Dbp2 binding sites.

Chromatin immunoprecipitation (ChIP)

ChIP was performed as described (Cloutier et al., 2013). Quantitative PCR was performed using Bio-Rad CFX96 Real-time system using PrimeTime Assay primers purchased from IDT (Table 2.5). Quantitative qPCR results from 3X-FLAG ChIP experiments were normalized using RNAPII ChIP from the same lysates to account for differences in transcriptional activity. The significance of the difference was tested using two-sample t-test assuming unequal variances. Strains used for ChIP analysis are listed in Table 2.1.

Preparation of Structure-seq libraries

The Structure-seq method was adapted from prior studies (Ding et al., 2015). 50 mL of yeast cells were grown in YP with 2% glucose at 30 °C to an OD_{600nm} of 0.5~0.7. Dimethyl sulfate (DMS)

was added to a final concentration of 10 mM and incubated for 10 min at 30 °C with vigorous shaking. The reaction was quenched with 75 mL of 4.8 M 2-mercaptoethanol (BME) and 25 mL of isoamyl alcohol. Cells were harvested by centrifugation and cell pellets were washed again with 4.8 M BME, followed by AE buffer (50 mM sodium acetate pH 5.2, 10 mM EDTA). Polyadenylated RNAs were purified and reverse transcribed by SuperScript III (ThermoFisher Scientific) as described (Ding et al., 2015). After extraction, the cDNAs were resolved by denaturing PAGE and visualized with SYBR Gold. cDNAs longer than 30 nt were isolated and eluted from the gel in TEN buffer (Ding et al., 2015) at 4 °C overnight. Gel purified cDNAs were ethanol-precipitated and re-suspended in water. An ssDNA linker (Table 2.6) was ligated to the cDNA 3' ends using CircLigase I (Epicentre) as described (Ding et al., 2015). After ligation, cDNAs above 60 nts were gel purified as above and subjected to PCR amplification as described (Ding et al., 2015). PCR products were resolved by denaturing PAGE, and products above 180 bp were gel purified. After elution, the library was ethanol precipitated and re-suspended in water. All the samples were analyzed by Agilent Bioanalyzer to determine the size distribution of the library. A total of 10 libraries, including three replicates of the wild type and two replicates of *dbp2Δ*, with or without DMS treatment, were sequenced on the Illumina HiSeq 2500 platform for 2 x 100 bp paired-end cycle run.

Processing of Structure-seq data

Illumina adaptors were removed using Trimmomatic (version 0.36) (Bolger et al., 2014). The random trimers were trimmed from the 5' end of forward reads using cutadapt (version 1.9.1) (Martin, 2011). The processed reads were mapped to the S288C reference genome (R64-2-1, from *Saccharomyces* Genome Database) using STAR (version 2.5.2b) and only uniquely mapped reads (MAPQ = 255 after STAR alignment) were kept for the subsequent analysis. The transcriptome annotation was as described in the 'Metagene Analysis' section. Ignoring genes with sequence overlaps with at least one other gene on the same strand, we retained 4681 mRNAs and 77 snoRNAs for differential DMS reactivity analysis. Reads were grouped according to their source RNA, and the start and end indices from genomic alignment of each read were converted to the RNA coordinates with the start of 5' UTR or mature 5' end as position +1.

Calculation of DMS reactivities for each transcript

The number of reads starting 1 nt downstream of each nucleotide were tallied to get the detection counts for the nucleotide. In addition, the number of reads starting -1 to +1 of each nucleotide, and ending anywhere downstream of the nucleotide were tallied as its “local coverage.” Detection rates were calculated as ratio of detection counts to local coverage for each nucleotide (Choudhary et al., 2016). Raw reactivities were calculated by combining the information from treated (+) and untreated (-) samples of the same biological replicate. For nucleotide i , raw reactivity, R_i , for a nucleotide i was obtained as

$$R_i = \max\left(\frac{r_i^+ - r_i^-}{1 - r_i^-}, 0\right),$$

where r_i^+ and r_i^- are detection rates at nucleotide i for treated and untreated samples, respectively (Aviran et al., 2011a, 2011b). Reactivities for Gs and Us were masked as missing information. Next, raw reactivities in each replicate were normalized using a 2-8% approach (Sloma and Mathews, 2015).

Differential analysis of DMS reactivity changes in the wild type and $dbp2\Delta$

Transcripts with average local coverage ≥ 250 were considered in the following analysis. We used a recently developed method, dStruct, for differential analysis of reactivities (Choudhary et al., 2019). dStruct compares inherent variation in biological replicates to *DBP2*-dependent variation at the transcript level. The former kind of variation was calculated using the three wild-type replicates and the latter was calculated using the two replicates of *dbp2* Δ strain and a wild type replicate performed in an independent batch. Results with both the raw and the FDR-adjusted p-values less than 0.05 were considered significant.

Identification of 3' extended mRNAs in $dbp2\Delta$

For this analysis, only transcripts without downstream overlapping genes in the sense direction and with more than 5 read counts were considered (3428 mRNAs). For each mRNA, reads from untreated (no DMS) Structure-seq libraries mapped to the ORF or 150 nt downstream of the 3' UTR (referred as extended region below) were counted using the summarizeOverlaps function (IntersectionNotEmpty mode) in the Bioconductor package “GenomicAlignments” (version 1.8.4). Only transcripts meeting the following conditions were analyzed: (1) the extended region does not

overlap with the downstream transcript, (2) with more than 0.97 counts per million and (3) detected in at least two libraries, were analyzed. The differential expression of ORFs and extended regions between wild type and *dbp2Δ* was analyzed using the Bioconductor edgeR package (version 3.14.0, (Robinson et al., 2010)). ORFs or extended regions with a false discovery rate (FDR) equal to or less than 0.05 were considered differentially expressed in *dbp2Δ* compared to the wild type. For each transcript, if the reads of the ORF were not changed or down-regulated, but the extended region was up-regulated in *dbp2Δ*, then it was regarded as a 3' extended mRNA in *dbp2Δ*. If the ORF was up-regulated in *dbp2Δ*, the extended region in *dbp2Δ* should be up-regulated and have a fold change twice higher than the fold change in the ORF to be considered as a 3' extended mRNA in *dbp2Δ* compared to the wild type.

Analysis of the relationship between DMS reactivity change and Dbp2 binding

Metagene plots in mRNAs were generated using custom R scripts. For each replicate, the 5' UTR, ORF, and 3' UTR, were divided into 10, 80, and 10 bins, respectively. DMS reactivities of As and Cs in each bin were tallied. The value in each bin of each transcript from biological replicates was averaged. For each bin, the values from all of the transcripts analyzed were averaged to represent the overall DMS reactivity. Values for all bins were then plotted as a line graph across the whole transcript. To visualize structural changes resulting from deletion of *DBP2*, reactivities in wild type were subtracted from reactivities in *dbp2Δ*. The metagene-based difference in reactivity was plotted as described above. For each nucleotide in each analyzed mRNA transcript, the distance between the nucleotide and the closest Dbp2 binding site was calculated, and the reactivity value was averaged across wild type or *dbp2Δ* biological replicates. The average reactivity at each distance was then calculated and the *DBP2*-dependent changes were derived by subtracting reactivities in *dbp2Δ* by those in wild type. The relationship was demonstrated by plotting the change of reactivities over a range of distance to Dbp2 binding sites.

Termination reporter assay

Yeast strains *cup1Δ* and *cup1Δdbp2Δ* cells were transformed with pGAC24, pGAC24-*CYC1 TER*, and pGAC24-*YOP1 TER*. Cells were grown in SC-LEU with 2% glucose and spotted in 5-fold serial dilutions on SC-LEU with 2% glucose plates containing 0 or 1.2 mM CuSO₄. Plates were incubated at 30 °C. For liquid growth assays, *cup1Δ* and *cup1Δdbp2Δ* cells containing pGAC24,

GAC24-*YOP1 TER*, pGAC24-*YOP1 ter-mut1* or pGAC24-*YOP1 ter-mut1+mut2* were incubated in SC-LEU with 2% glucose media containing 0 or 0.4 mM CuSO₄ at 30 °C with shaking in a microplate reader (Biotek epoch2, Winooski, VT). OD_{600nm} values were taken every 30 min. The plot was produced with the average of three replicates.

Data availability

Strains and plasmids are available upon request. Supplementary tables that are in excel format are uploaded to figshare. Raw reads of iCLIP-seq and structure-seq and processed data are deposited on GEO with the accession number GSE106479. Scripts for bioinformatics analysis are deposited on Github: https://github.com/karenlai0222/Dbp2_iCLIP_Structure.

2.3 Results

2.3.1 Dbp2 associates predominantly with snoRNAs and mRNAs

Dbp2 has been shown to function in mRNA metabolism and alter gene expression (Barta and Iggo, 1995; Beck et al., 2014; Cloutier et al., 2012; Ma et al., 2013; Wang et al., 2017). Consistently, in a very recent study, Dbp2 was reported to associate with coding and non-coding RNAs, including RNAPII-transcribed small nucleolar RNAs (snoRNAs) and mRNAs (Tedeschi et al., 2018). To gain insight into the precise role of Dbp2 on snoRNAs and mRNAs, we comprehensively identified all RNA targets bound by Dbp2 *in vivo*. In contrast to prior studies, which utilized a denaturing protocol in conjunction with individual-nucleotide resolution crosslinking and immunoprecipitation (iCLIP), we used a more standard native iCLIP-seq called FAST-iCLIP (Flynn et al., 2015) using an endogenous, C-terminally 3XFLAG-tagged *DBP2* strain. This method resulted in a significantly greater yield, with an average of ~1.2 million uniquely mapped reads per Dbp2 replicate across three biological replicates.

We then determined binding sites from reverse transcription stops induced at the site of crosslinking (König et al., 2010) and the binding site counts for each RNA class, which were highly reproducible across replicates (Figure 2.1). This resulted in a distribution of uniquely mapped reads among RNA classes similar to what was derived from the uniquely mapped reads in the prior study (data not shown) (Tedeschi et al., 2018). Among all unique mapped reads in Dbp2-iCLIP, equal proportions (37%) mapped to snoRNAs and mRNAs, and to a lesser content to other classes

(Figure 2.2A). The vast majority of transcribed snoRNAs and approximately 1/3 of protein-coding transcripts were isolated as Dbp2-bound targets. This is consistent with prior studies showing misregulation of both snoRNAs and protein-coding genes in the absence of *DBP2* (Cloutier et al., 2012; Ma et al., 2016; Beck et al., 2014).

2.3.2 Dbp2 promotes efficient transcription termination of snoRNAs

Dbp2 interacts physically with Sen1 (Tedeschi et al., 2018), an RNA-DNA helicase involved in transcriptional termination of snoRNAs in conjunction with RNA-binding proteins Nrd1 and Nab3 (Arndt and Reines, 2015). To determine if Dbp2 plays a role in snoRNA termination, we performed RNAPII chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-seq) for both wild-type and *dbp2Δ* cells expressing an endogenous, C-terminally 3XFLAG-tagged Rpb3 subunit. Input and immunoprecipitation (IP) libraries were prepared from 3 biological replicates of both wild-type and *dbp2Δ* cells and resulted in an average of ~1.8 million reads per replicate. We then determined the distribution of RNAPII along genes using a package for ChIP-seq data analysis, MACS2 (Zhang et al., 2008), and asked if this occupancy is altered in *dbp2Δ* at snoRNA transcriptional termination sites (Schaughency et al., 2014). In wild type cells, RNAPII shows an average occupancy profile across all monocistronic snoRNA genes that conforms to a bell shaped curve with a peak near -200 bp, with respect to the termination site (0 bp) (Schaughency et al., 2014), that corresponds to the transcription start site (Figure 2.2B). In *dbp2Δ* cells, this peak is slightly reduced in height and shifted ~20 bp downstream. Moreover, the average RNAPII occupancy remains higher after the annotated termination site and shows a less steep reduction in *dbp2Δ* cells compared to wild type. This pattern is highly similar to the RNAPII ChIP profiles following nuclear depletion of either Sen1 or Nrd1 (Schaughency et al., 2014). It has been proposed that during termination of snoRNA transcription, Nrd1 and Nab3 recognize sequence motifs in the precursor RNA and recruit Sen1, leading to termination of RNAPII elongation (Arndt and Reines, 2015). Inspection of individual snoRNA gene profiles also reveals RNAPII accumulation in the absence of *DBP2* after the annotated mature 3' ends, across regions that correspond to previously identified binding sites for Nrd1, Nab3, and at the termination sites of precursor snoRNA transcripts (Figure 2.2C) (Jamonnak et al., 2011). Interestingly, and consistent with prior studies (Tedeschi et al., 2018), Dbp2 iCLIP sites correspond to Nrd1, Nab3 and Sen1-binding sites at some snoRNAs in the mature transcript region

(SNR3, Figure 2.2C) but not others (SNR189 and SNR46). This binding pattern suggests that Dbp2 also binds to the mature, fully processed snoRNAs.

Because snoRNA gene termination is coupled to 3' end processing, we then analyzed our Dbp2-iCLIP reads to detect the presence of unprocessed 3' extensions in the Dbp2 RNA-binding data. This revealed sequences that mapped to 11 different snoRNA species that also contained 1 to 30 unprocessed nucleotides followed by short stretches of non-templated As on the 3' ends (Figure 2.2D). These sequences likely correspond to processing intermediates of the TRAMP complex and nuclear exosome, which are coupled with NNS-dependent termination (Arndt and Reines, 2015). Similar processing intermediates have been observed in PAR-CLIP data of NNS factors (Jamonnak et al., 2011) and upon analysis of snoRNAs in *rrp6Δ* mutants (Grzechnik and Kufel, 2008). Taken together, this suggests that Dbp2 functions in termination and maturation of snoRNAs.

2.3.3 Dbp2 shares RNA-binding profiles with Nrd1 and Nab3 on protein-coding transcripts and promotes loading of Nrd1 on gene 5' Ends

Loss of *DBP2* results in accumulation of a bicistronic GAL10-GAL7 transcript (Cloutier et al., 2012), suggesting that Dbp2 may also function in termination of protein-coding genes in addition to snoRNAs. To gain insight into the role of Dbp2 binding on protein coding genes, we then analyzed the binding pattern of Dbp2 on protein-coding transcripts using our iCLIP data set. Metagene analysis revealed Dbp2 binding across the entire length of targeted mRNAs, with highest accumulations at the 5' end of the ORF and the 3'UTR (Figure 2.3A, red line). To determine if this mRNA-binding profile is similar to NNS components, we overlapped our meta-analysis of Dbp2 binding with meta-analysis profiles of Nrd1, Nab3, and Sen1 generated from previously published datasets (Creamer et al., 2011). This revealed a strikingly similar pattern of Nrd1 and Nab3 binding to Dbp2, with all three proteins exhibiting peaks at 5' and 3' ends of the genes (Figure 2.3A). Analyzing the distribution of Nrd1 and Nab3 binding sites relative to the Dbp2 crosslink site on mRNAs revealed a broad peak of Nab3 and Nrd1 binding at or near Dbp2 (within ± 20 nt) (Figure 2.3B), suggestive of overlapping binding for all three factors. Interestingly, we also observed a periodicity in Nrd1 accumulation of approximately 25 nts within the 160 nt window, with decreased levels of Nrd1 accumulation from -55 to +55 with respect to the Dbp2 crosslinking site (Figure 2.3B). This pattern was not observed for Nab3. We then analyzed Dbp2 iCLIP reads for the presence of enriched sequence motifs using HOMER (Heinz et al., 2010).

Despite the fact that the vast majority of DEAD-box RNA helicases exhibit sequence-independent RNA-binding *in vitro* (Gilman et al., 2017), we found three, significantly enriched motifs in Dbp2 bound 3' UTRs (Figure 2.3C). The first motif with the highest enrichment (Figure 2.3C, motif 1) is strikingly similar to the UGUA Nrd1 RNA-binding motif (Figure 2.3C, top right, (Creamer et al., 2011)). The second most enriched motif is U/C rich (Figure 2.3C, motif 2) and somewhat similar to the Nab3 sequence motif (Figure 2.3C, bottom right, (Creamer et al., 2011)). Motif 2 is also similar to the U/C rich motif of Rna15 (UUUUCUU, (Baejen et al., 2014)), a component of the cleavage and polyadenylation factor (CPF) complex, that plays the predominant role in termination of protein-coding genes (Mischo and Proudfoot, 2013). However, no global similarity was seen between the meta-analysis profiles of Rna15 and Dbp2 (Figure 2.4A), nor do the two proteins share common occupancy sites when averaged across all Dbp2-bound mRNAs (Figure 2.4B).

The common binding profiles of Dbp2 with Nrd1 and Nab3 suggest that this enzyme may function in concert with the NNS complex on protein-coding genes in addition to snoRNA genes. Association of the NNS complex with the 5' ends of protein-coding genes has been associated with transcription attenuation (Arigo et al., 2006; Kim and Levin, 2011; Kuehner and Brow, 2008). Interestingly, comparison of individual RNA-binding tracks on known NNS-targeted protein-coding transcripts, *NRD1* and *PCF11* (Creamer et al., 2011), revealed strikingly similar binding patterns of Nrd1, Nab3, and Dbp2 at the 5' ends of both genes (Figure 2.3D). We then asked if *DBP2* is necessary for Nrd1 binding to transcribed genes by conducting ChIP of a *3XFLAG*-tagged Nrd1 protein in wild type and *dbp2Δ* cells. This revealed an approximately 50% decrease in Nrd1 association at the 5' end of both *NRD1* and *PCF11* genes in the absence of *DBP2* (Figure 2.3E). A western blot shows that decreased Nrd1 binding in *dbp2Δ* cells is not due to decreased Nrd1 protein abundance (Figure 2.3F). Taken together, these results suggest that Dbp2 may function in concert with the NNS complex at targeted protein-coding genes, possibly by promoting binding of Nrd1 to nascent RNA.

2.3.4 *DBP2* facilitates proper transcription termination of a subset of mRNAs

Next, we asked if RNAPII accumulates downstream of the annotated termination site in protein-coding genes in *dbp2Δ* cells similar to snoRNA genes above. Importantly, this analysis was restricted to exclude overlapping genes encoded in tandem to minimize read assignment

ambiguity. We also performed RNA-sequencing (RNA-seq) of poly(A)-selected RNAs from wild type and *dbp2Δ* cells to identify gene products with putative termination defects as evidenced by 3' extensions. Protein-coding transcripts with 3' extensions were identified from the RNA-seq by the presence of reads mapping greater than 150 bp downstream of the annotated 3'UTR (Nagalakshmi et al., 2008; Yassour et al., 2009). This arbitrary definition was selected as 90% of RNAPII termination events occur within a 50 bp window after the polyadenylation site (Baejen et al., 2017).

Whereas RNAPII exhibited a broad peak of maximal occupancy centered at 30 bp downstream of polyadenylation site (annotated as 0 bp) in wild type cells, loss of *DBP2* resulted in a slight shift downstream with the highest peak central at 60 bp downstream (Figure 2.5A). We also observed an unexpected reduction of occupancy 200 bp and further to either side of the polyadenylation site in *dbp2Δ* cells.

Next, we analyzed our RNA-seq data for the presence of 3' extended mRNAs. This yielded 824 protein-coding transcripts, corresponding to ~14% protein-coding genes, with 3' extensions in *dbp2Δ* cells (Table 2.7). (Note that this number is likely an under-representation due to the criteria outlined above for identification of 3' extended gene products.) A Fisher's exact test revealed a statistically significant overlap between mRNAs bound by Dbp2 at their 3' ends and those 3' extended isoforms, suggesting a correlation between Dbp2 binding and suppression of 3' extension (Figure 2.5B and Table 2.8). The fact that not all aberrant transcripts in *dbp2Δ* cells are represented in our Dbp2 iCLIP is not surprising, as iCLIP isolation depends on both RNA-binding and RNA sequence context (Sugimoto et al., 2012).

By calculating the percentage of mRNAs with 3' extensions versus total mRNAs for a given gene product, using the following formula: read counts that map 150 bp downstream of polyadenylation site/read counts mapped to the ORF, we found that the percentage of 3' extended mRNAs ranged from 0.0 to 18.0% in the wild type (median: 2.3%) and from 0.1 to 55.0% in *dbp2Δ* (median: 8.4%) (Figure 2.5C-D). The presence of 3' extended products in wild type cells is consistent with the recently described, heterogeneous nature of termination site choice in *S. cerevisiae* (Moqtaderi et al., 2018). Loss of *DBP2* appears to increase the proportion of 3' extended products per gene, as illustrated most clearly by the box plot distribution (Figure 2.5D). This increased abundance could be due to read-through of a predominant termination site, increased stability of the 3' extended mRNAs in the absence of *dbp2Δ*, as recently described for mRNA 3'

isoforms (Moqtaderi et al., 2018), or a combination of both. To test the possibility of long 3' isoform accumulation in *dbp2Δ*, we compared the end positions in our analysis for read-through with the most 3' polyadenylation sites (Pelechano et al., 2013). We found that the majority (~77%) of the identified, extended transcripts had reads extended past the most 3' polyadenylation site, suggestive of read-through products, whereas the other extended transcripts can be longer isoforms accumulated in *dbp2Δ*. Future study with 3' end-specific sequencing would be needed to determine the precise *DBP2*-dependent change in polyadenylation site usage. We then analyzed RNAPII occupancy for the 824 genes with 3' extended mRNAs in *dbp2Δ* cells. This revealed accumulation of RNAPII in *dbp2Δ* across a 400-bp window centered at the polyadenylation site with the most striking accumulation 0-150 bp downstream (Figure 2.5E). This accumulation is centered 50 bp downstream in *dbp2Δ* cells as compared to wild type, and is reminiscent of nuclear depletion of Rat1 or Ysh1, nucleases required for efficient RNAPII termination (Baejen et al., 2017; Schaughency et al., 2014). In addition to genes with delayed RNAPII pausing or termination, we did observe that, in a small subset of genes (~250), the most 3' peak summit of RNAPII in *dbp2Δ* is slightly upstream compared to the peak in the wild type. This may contribute to the reduction of RNAPII signal at > 200 bp downstream of polyadenylation sites in *dbp2Δ* observed in Figure 2.5A. However, the resolution of this analysis is not sufficient to determine the exact mechanism contributing to the shift of peak toward upstream and will need further characterization in the future.

2.3.5 Dbp2 facilitates loading of termination factors at 3' ends of protein-coding genes

To gain insight into the mechanism for Dbp2-dependent termination of protein-coding transcripts, we then inspected the RNAPII occupancy of three genes that produce 3' extended mRNAs in *dbp2Δ* that are also bound by Dbp2 in their 3' UTRs in wild type cells. These three genes are: *YOP1*, an intron-containing, shorter gene with an average length 3'UTR (675 bp ORF, 140 bp 3'UTR), *RBG1*, an intronless, longer gene with an average 3'UTR (1100 bp ORF, 167 bp 3'UTR), and *YNL190W*, a shorter gene with a long 3' UTR (614 bp ORF, 543 bp 3'UTR). Note that the average gene length and 3'UTR for all protein-coding genes are 1343 and 188 bp, respectively. All three profiles showed accumulation of RNAPII that extends further downstream in *dbp2Δ* than in wild type cells, indicative of defective termination (Figure 2.6A). Interestingly, we also observed a decrease in RNAPII accumulation at the 5' ends of all three genes, which likely

explains the decreased accumulation towards the 5' end of genes in our combined RNAPII profile (Figure 2.5A). A similar decrease was also noted for *sen1* mutants on short, protein-coding genes and was speculated to be due to decreased initiation (Steinmetz et al., 2006). However, we observed no difference in the abundance between wild type and *dbp2Δ* cells of transcripts of these three genes. Moreover, *YNL190W* shows a change in the overall pattern of RNAPII accumulation. This suggests that Dbp2 may also function in the rate of RNAPII progression along a given gene. Previously, the NNS complex was shown to bind the 3' UTRs of a subset of mRNAs (Creamer et al., 2011; Webb et al., 2014); however, the precise functional role of this complex in 3'UTRs was not determined. Comparison of the RNA-binding profiles of Dbp2, Nrd1, Nab3, and Sen1 revealed that Dbp2 binding in the 3'UTRs of these transcripts overlaps with the binding of one or more factors in the NNS complex (Figure 2.6D-F). Since *DBP2* promotes efficient recruitment of Nrd1 to the 5' ends of NNS-targeted protein coding genes (Figure 2.3E), we then asked if this was also the case at gene 3' ends by conducting ChIP of Nrd1-3XFLAG at the 5' end (1), 3'UTR (2), and downstream of the polyadenylation site (3) (Figure 2.6C). Note that we were unable to assay 3'UTR (2) binding to *RBG1* due to technical reasons (lack of efficient primer-binding sites for qPCR because of AT-richness). Similar to the requirement for *DBP2* at the 5' ends of known NNS targeted genes, we also observed reduced Nrd1 association at both the 5' and 3' ends of our 3 candidate genes in the absence of *DBP2*, regardless of the presence of an intron or length of the 3'UTR (Figure 2.6C). Interestingly, loss of *DBP2* also decreased the association of a 3X-FLAG-tagged Rna15, but this decrease was only statistically significant downstream of the polyadenylation site, in line with the primary location of RNA-binding by this protein after the 3'UTR (Baejen et al., 2014) (Figure 2.6B, bottom panels). This suggests that Dbp2 may also promote efficient association of Rna15 on protein-coding genes, despite the distinct RNA-binding patterns of these two proteins (Figure 2.6B, 2.4A-B). This may be due to an as-of-yet unidentified role for Dbp2 in CPF-dependent termination or to an indirect effect of reduced Nrd1 binding at sites within the 3'UTR.

2.3.6 Dbp2 modulates RNA/ribonucleoprotein (RNP) structures *in vivo*

Given that Dbp2 is an efficient helicase *in vitro* (Ma et al., 2013), and DEAD-box proteins have been shown to remodel RNA structures *in vivo* (Guenther et al., 2018), we hypothesized that Dbp2 might remodel RNA structures and/or RNA-protein complexes of RNAPII transcripts to

promote termination and subsequent mRNA metabolism steps. This is in line with the fact that sequestration of polyadenylation signal sequences in secondary structures can impede CPF-dependent termination (Chen and Wilusz, 1998; Klasens et al., 1998). This is likely to be true for NNS binding as both Nrd1 and Nab3 recognize RNA motifs in the context of single stranded RNA (Arndt and Reines, 2015; Singh et al., 2007). Moreover, *DBP2* is required for association of mRNA export factors Yra1 and Mex67 (Ma et al., 2013), which act after termination. Moreover, sequestration of polyadenylation signal sequences in secondary structures can impede CPF-dependent termination (Chen and Wilusz, 1998; Klasens et al., 1998). This is likely to be true for NNS binding as both Nrd1 and Nab3 recognize RNA motifs in the context of single stranded RNA (Arndt and Reines, 2015; Singh et al., 2007).

To test this, we conducted Structure-seq of total, poly(A)-selected RNA in both wild type and *dbp2Δ* cells. Structure-seq uses the cell permeable compound DMS to preferentially methylate adenosine and cytosine residues that are not involved in Watson-Crick base pairing (Peattie and Gilbert, 1980) or not protected by proteins, combined with next-generation sequencing to observe RNA structure and compositional changes genome-wide (Ding et al., 2015). Methylated nucleotides were detected as reverse transcription “stops”, after library construction and RNA sequencing, and then translated into nucleotide-level reactivities to DMS and a prediction of increased or decreased protection between wild type and *dbp2Δ* (see Materials and Methods). Total poly(A)+ RNA was used for Structure-seq to enable direct comparison with 3' extended transcripts in our RNA-seq and because there is currently no method available to map secondary structures in nascent RNA transcripts in eukaryotes. Consistent with prior studies, we observed RT-stops indicative of DMS-dependent methylation predominantly at As and, to a lesser extent, at Cs in DMS-treated samples compared to untreated (Figure 2.7) (Ding et al., 2014; Rouskin et al., 2013). We then analyzed differential DMS reactivity in wild type and *dbp2Δ* for each transcript using a newly developed method, dStruct, for Structure-seq analysis (Choudhary et al., 2019). It identifies transcripts and regions that manifest significantly more variation between strains than among samples of the same strain.

Our analysis identified 612 protein-coding transcripts with significant, *DBP2*-dependent changes in DMS reactivity (p-value < 0.05, false discovery rate (FDR) < 0.05, Table 2.9). Importantly, these transcripts were significantly enriched in Dbp2-bound transcripts identified in our iCLIP-seq (p-value < 2.2e-16, Table 2.10), suggesting that these changes are directly due to

Dbp2 activity. A metagene analysis of differential reactivity of these 612 transcripts revealed a lower overall DMS reactivity indicative of increased nucleotide protection in *dbp2Δ* as compared to wild type, centered at the Dbp2-binding site (Figure 2.8A, left). This trend was also observed when the analysis was restricted to 3' UTRs (Figure 2.8A, right). Interestingly, the *DBP2*-dependent decrease in DMS accessibility spans ~60-70 nts in length around the Dbp2-binding site, suggestive of a broad region of Dbp2-dependent RNA/RNP structural remodeling. This pattern is unlikely to be due to Dbp2-binding alone as DEAD-box proteins have a binding-site of approximately 6 nts along the sugar-phosphate backbone (Andersen et al., 2006). Moreover, one would expect decreased reactivity in wild type cells rather than *dbp2Δ* cells, if this pattern was due to Dbp2 binding. Instead, this is more consistent with a region of increased RNA structures and/or RNP composition.

Next, we inspected the reactivity profiles of 3' UTRs of *YOP1*, *RBG1*, and *YNL190W* (Figure 2.8B-D). Whereas both increased and decreased reactivity was observed across all three selected 3'UTRs (Figure 2.8B-D, top and middle panels), nucleotides exhibiting decreased reactivity in the absence of *DBP2* gave the largest difference (Figure 2.8B-D, middle panels). Interestingly, these sites appear to correspond to those that exhibit the least protection (most reactive) in wild type cells in these three representative transcripts. This indicates that, while the overall 3'UTR structure, which can include both base-pairing and protein-binding, is changed in the absence of *DBP2*, there may be a tendency for regions that are typically unprotected to be protected. Moreover, Dbp2-binding sites are found within close proximity of these regions (Figure 2.8B). This is consistent with RNA/RNP structural remodeling of 3'UTRs by Dbp2. While we were able to analyze some snoRNAs for *DBP2*-dependent structural changes (Figure 2.9), we were unable to analyze regions of termination/processing due to insufficient coverage for dStruct analysis in preprocessed precursor snoRNAs. Regardless, our results above, and the known enzymatic role of Dbp2 as a DEAD-box RNA helicase, suggest that this enzyme may remodel RNA structure during termination for multiple RNAPII transcripts.

2.3.7 *DBP2*-dependent RNA structure correlates with *DBP2*-dependent termination

To understand the relationship between binding and structural remodeling of 3'UTRs of protein-coding transcripts with efficient termination, we qualitatively compared differential reactivity profiles across different datasets. First, we compared the DMS reactivities between wild

type and *dbp2Δ* cells by plotting the difference in reactivity (*dbp2Δ* minus wild type) for all transcripts that are bound or not bound by Dbp2. This revealed reduced reactivity across Dbp2-bound transcripts in the absence of *DBP2* across the length of the 5'UTR, ORF, and 3'UTR (Figure 2.10A). Second, we compared differential reactivities of transcripts that exhibit 3' extensions in *dbp2Δ* to those that do not (Figure 2.10B). This revealed that transcripts from genes with *DBP2*-dependent termination are more protected across their 3'UTRs in the absence of *DBP2* (Figure 2.10B). Consistent with the latter, a Fisher's exact test showed significant correlation between the presence of *DBP2*-dependent differential reactivity and *DBP2*-dependent termination (p-value = 9.4×10^{-16} , Figure 2.10C & Table 2.11). Moreover, meta-analysis of Dbp2 binding near mRNA 3' ends revealed a sharp peak at the immediate upstream of the polyadenylation site specifically in the group of mRNAs with 3' extension in *dbp2Δ* as compared to those without 3' extension (Figure 2.10D), in line with the pattern of structural change observed in Figure 2.10B. This suggests the Dbp2-dependent remodeling of structures and/or proteins within the 3'UTR as a likely mechanism for *DBP2*-dependent reactivity.

To test if the presence and stability of secondary structures in 3' UTRs dictates a requirement for Dbp2 in termination, we utilized a transcriptional termination reporter growth assay that previously enabled identification of both *cis*-acting termination elements and *trans*-acting termination factors (Steinmetz and Brow, 2003; Steinmetz et al., 2001). Briefly, the reporter encodes a chimeric actin-metallothionein (*ACT1-CUP1*) transcript that includes the intron of *ACT1* (Figure 2.11A). Insertion of a terminator within the intron prevents expression of the *ACT1-CUP1* chimera and results in death of cells lacking endogenous *CUP1* (*cup1Δ* cells) on copper-containing plates. Defective termination, however, allows read-through of the internal terminator and resistant growth on different concentrations of copper, the latter of which is inversely proportional to the efficiency of termination (Steinmetz and Brow, 2003).

First, we generated a putative, *DBP2*-dependent termination reporter by inserting the 3'UTR region of *YOP1* into the intron of the *ACT1-CUP1* reporter (Figure 2.11A). To determine if the reporter recapitulates the *DBP2*-dependent termination defects seen at the endogenous loci, strains lacking *CUP1* were first transformed with reporters either lacking a terminator (no terminator), containing an inserted *CYC1* terminator (*CYC1 TER*), or containing the *YOP1* 3'UTR. *CYC1* termination is independent of *DBP2* and serves as a negative control (Table 2.7). Without an internal terminator, the transcription of the reporter is terminated only at the end of *CUP1* gene

(Figure 2.11A), allowing production of the Act1-Cup1 chimera and enabling growth of both *cup1Δ* and *cup1Δ dbp2Δ* cells on copper (Figure 2.11B, no terminator). Insertion of *CYC1 TER* resulted in a copper-sensitive phenotype in both *cup1Δ* and *cup1Δ dbp2Δ*, consistent with termination in both strains prior to *CUP1* (Figure 2.11B, + *CYC1 TER*). In contrast, insertion of the *YOP1 3' UTR* prevented copper-resistant growth of *cup1Δ* cells but not *cup1Δ dbp2Δ* cells, indicating that the 3'UTR of *YOP1* is sufficient to elicit termination in wild type cells and that this termination is *DBP2*-dependent (Figure 2.11B, + *YOP1 3'UTR*).

We then generated a structure prediction of the 3'UTR of *YOP1* in the absence of *DBP2* using DMS reactivity data and the ViennaRNA package 2.0 (Lorenz et al., 2011) (Figure 2.11C). Interestingly, UA/UG-rich (Figure 2.11B, boxes) motifs, which are sequences recognized by Nrd1 and Nab3 (Schaughency et al., 2014), are present in predicted secondary structures. Based on the PAR-CLIP data of NNS factors (Creamer et al., 2011), Nrd1 and Nab3 crosslink to nucleotides 848 and 849, respectively, at the base of the 814 – 851 nt stem structure, whereas Sen1 crosslinks to nucleotides 773, 774, and 776 at the base of the 773 – 804 stem structure located upstream.

To test if the termination defect in *dbp2Δ* cells is dependent on the stability of RNA secondary structures within *YOP1 3'UTR*, we mutated the sequences within the reporter to destabilize one or both structures (Figure 2.11C-D). We then assayed growth in liquid culture over time to increase the sensitivity of the assay. Importantly, neither mutation enabled growth of *cup1Δ* cells in the presence of copper, indicating that these mutations do not alter termination efficiency in wild type cells (Figure 2.11D, top panels). Consistent with our plate assay, cells lacking *DBP2* exhibited copper-resistant growth with the *YOP1 3'UTR* reporter. However, this resistance was reduced in both reporter mutants, indicative of increased termination (Figure 2.11D, bottom panels). Furthermore, the amount of secondary structure in the 3' UTR correlated with the requirement for *DBP2* in termination as evidenced by decreased growth of the *cup1Δ dbp2Δ* cells with the mut1+2 reporter over mut1 alone, indicating that the secondary structure of a given 3'UTR determines the requirement for Dbp2 in termination. This suggests that Dbp2 may promote termination efficiency by remodeling RNA/RNP structure in certain protein-coding genes. Taken together, our data point to a model whereby Dbp2 promotes efficient termination of RNPII transcription by modulating RNA/RNP structures to facilitate recruitment of NNS components. Future studies are necessary to

determine the precise role of NNS-complex members at the 3' ends of protein-coding genes (see Discussion).

2.4 Discussion

Termination of RNAPII in *S. cerevisiae* is executed predominantly by the CPF or NNS complex, which function on protein-coding genes or non-coding RNAs, respectively. This distribution of tasks allows coupling of termination with 3' processing steps that are appropriate for the given transcript, such as endonucleolytic cleavage and addition of a poly(A) tail to maturing mRNAs or 3' end trimming or decay of non-coding RNAs. While the histone modification pattern of the transcribed gene, the phosphorylation status of RNAPII, and the sequence of the nascent transcript itself all dictate the mode and efficiency of termination (Proudfoot, 2016), the contribution of RNA structure and role of RNA helicases in this process has not been fully explored. By aggregating the results of multiple genome-wide analyses, we show that the RNA helicase Dbp2 promotes transcriptional termination and provide evidence that this mechanism likely involves remodeling of nascent transcripts and recruitment of Nrd1 to targeted genes.

In budding yeast, NNS-dependent termination of non-coding genes is coupled to 3' end processing of nascent transcripts by nuclear exosome with the aid of the Trf4-Air2-Mtr4 Polyadenylation (TRAMP) complex, which adds short stretches of oligo(A) to the 3' ends of NNS products to promote 3' end maturation (Arndt and Reines, 2015). The Sen1 RNA-DNA helicase, a component of the NNS complex, is thought to facilitate termination by unwinding the RNA-DNA hybrid produced in the wake of RNAPII (Conti et al., 2017; Han et al., 2017). Our results show that loss of *DBP2* results in a shift of RNAPII accumulation downstream of annotated snoRNA termination sites, well characterized targets of the NNS complex, indicative of a role for Dbp2 in NNS-dependent termination. Dbp2 may also facilitate termination-coupled processing steps, as evidenced by the isolation of Dbp2-bound snoRNA processing intermediates with non-templated A's. These processing intermediates are highly reminiscent of semi-mature snoRNA species produced by the distributive polyadenylation polymerase activity of TRAMP in conjunction with Rrp6-dependent processing (Grzechnik and Kufel, 2008) and have also been isolated by Nrd1 PAR-CLIP studies (Jamonnak et al., 2011). This role is also supported by synthetic lethal interactions between *RRP6* and *AIR2* with *DBP2* (Cloutier et al., 2012; Wilmes et al., 2008) and the fact that loss of *DBP2* results in upregulation and 3' extension of the GAL10s

lncRNA (Cloutier et al., 2012). It is interesting that Dbp2 accumulates across the snoRNA body while promoting termination (Figure 2.2B-C). A likely explanation is that Dbp2 also functions in snoRNA-guided rRNA modification, a role that is supported by the nucleolar accumulation of Dbp2 and its association with ribosome subunits (Cloutier et al., 2012; Martin et al., 2013).

In addition to snoRNAs, our results also define a role for Dbp2 in termination of protein-coding genes. Whereas the predominant mechanism for termination of protein-coding genes involves the CPF complex, our results suggest that, like snoRNA genes, Dbp2-dependent termination of protein-coding genes may also involve the NNS complex. This is evidenced by enrichment of Nrd1 and, to a lesser extent, Nab3 consensus motifs as Dbp2-bound RNA targets (Figure 2.3C), *DBP2*-dependent binding of Nrd1 to the 5' ends and 3' UTRs of targeted protein coding genes (Figure 2.3E & 2.6C), and the recapitulation of *DBP2*-dependent termination upon insertion of a Nrd1-targeted 3'UTR into a reporter construct (Figure 2.11). It should be noted, however, that while our results say that Dbp2-dependent termination is mediated through Nrd1, this does not mean that this mechanism is CPF-independent.

The decision of whether to terminate transcription using CPF or NNS is largely determined by the phosphorylation state of the C-terminal domain (CTD) of RNAPII, which is recognized by CTD-interacting domains (CID) in either Pcf11 or Nrd1, respectively (Chinchilla et al., 2012; Eick and Geyer, 2013; Noble et al., 2005). Specific phosphorylation patterns within the heptad repeat of the CTD mark the position of RNAPII during transcription, with high levels of serine 5 (Ser5) phosphorylation transitioning to serine 2 (Ser2) phosphorylation along the length of the gene (Harlen and Churchman, 2017). Nrd1 associates with Ser5 phosphorylated CTD while Pcf11 associates with Ser2, confining Nrd1 to termination of short transcripts and Pcf11 to longer, protein-coding ones (Chinchilla et al., 2012; Eick and Geyer, 2013; Noble et al., 2005). However, several genome-wide studies have revealed association of Nrd1 and Nab3 with the 3' ends of mRNA transcripts, in addition to canonical snoRNAs and CUTs (Creamer et al., 2011; Webb et al., 2014; Wlotzka et al., 2011). In some cases, this association serves as a form of quality control, with NNS-directed termination serving to direct exosome recruitment to mis-processed RNAs (Gudipati et al., 2012). In line with this is the observation that NNS-targeted protein-coding transcripts are frequently upregulated upon mutation of *nrd1* or components of the exosome (Gudipati et al., 2012; Webb et al., 2014). However, we observed no significant correlation between transcript upregulation and 3' extension in *dbp2Δ* cells (data not shown), suggesting that

DBP2-dependent termination of protein-coding genes may not be coupled with decay. Interestingly, studies have shown that the Pcf11 component of the CPF complex depends on initial binding and subsequent exchange with Nrd1 (Grzechnik et al., 2015). While this Nrd1-Pcf11 “swap” was documented at the 5’ ends of genes, consistent with the role of NNS on shorter transcripts, evidence suggests that the converse may also occur. The most striking example is the characterization of a failsafe termination mechanism for RNAPII complexes that involves Nrd1/Nab3 sites that are downstream of a polyadenylation site (Rondón et al., 2009). Interestingly, characterization of this form of failsafe termination utilized the *GAL10-GAL7* genes and associated intergenic region, which corresponds to the same region where *DBP2*-dependent termination of both protein-coding and non-coding genes was first described (Cloutier et al., 2012). Moreover, mounting evidence has found that the role of CTD phosphorylation in downstream processing steps may be more nuanced than the prevailing Ser5-Ser2 gradient model. While the latter does hold true for the vast majority of genes in *S. cerevisiae* (Suh et al., 2016) gene-specific phosphorylation patterns and the link between Ser7 phosphorylation and Nrd1 recruitment have also been described (Kim et al., 2010; Kubicek et al., 2012; Vasiljeva et al., 2008). One interesting possibility for the role of Dbp2 stems from the observation that the Nrd1 ortholog in fission yeast, Seb1, promotes pausing of RNAPII, in addition to termination site selection and 3’ end processing (Lemay et al., 2016; Liu et al., 2017; Parsa et al., 2018). While *S. pombe* does not use an NNS-like mechanism for termination of protein-coding or non-coding RNAs, with Seb1 recruiting CPF machinery to both types of RNAPII-transcribed genes (Larochelle et al., 2018), the fact that loss of *DBP2* results in a change in RNAPII accumulation at Nrd1-binding sites on individual genes (Figure 2.3 & 2.6) that is strikingly similar to *seb1* mutation suggests that termination defects described here may be coupled to RNAPII progression. Future studies are necessary to determine if *DBP2* influences RNAPII kinetics and if the kinetics are affected by aberrant RNA structure or reduced binding of Nrd1 in its absence.

Thirty years ago to date, the first DEAD-box RNA helicase was shown to have RNA duplex unwinding activity *in vitro* (Abramson et al., 1987; Hirling et al., 1989). Since this discovery, elegant biochemical and biophysical studies have provided an in-depth knowledge of how these enzymes function *in vitro* (Rudolph and Klostermeier, 2015); however, their precise functions *in vivo* have remained elusive. Recently, however, the Jankowsky lab capitalized on the rapidly advancing technologies to view RNA-binding sites and RNA structure/composition genome-wide,

providing the first evidence for widespread RNA remodeling by a DEAD-box RNA helicase that directly influences gene expression steps (Guenther et al., 2018). While those studies provide a link between the DEAD-box RNA helicase Ded1 and translation, results herein now extend the generality of RNA remodeling events catalyzed by these enzymes to nuclear gene expression steps. Loss of *DBP2* results in decreased DMS reactivity, mostly in 3' UTRs (Figure 2.8 & 2.10). Therefore, RNA/RNP structures may contribute to regulation of specific processes. It is likely that the altered mechanism of termination and extended length of the transcripts produced in *dbp2Δ* cells result in different compositions and/or stoichiometries of RBPs that account for the 40-60 nt stretch of reduced DMS reactivity (Figure 2.8A). It has been established that efficient termination is a pre-requisite for recruitment of mRNA export factors to mature mRNAs (Dunn et al., 2005; Hammell et al., 2002; Kessler et al., 1997) and our laboratory has shown that loss of *DBP2* reduces association of the nuclear poly(A)-binding protein, Nab2, and mRNA export factor, Mex67 on poly(A)+ RNAs (Ma et al., 2013). Moreover, recent studies by the Struhl laboratory found that the protein composition of the 3' isoform produced is influenced by the length of the transcript (Moqtaderi et al., 2018). However, the fact that the 40-60 nt stretch is centered at the Dbp2-binding site suggests that the altered folding and/or composition is directly due to loss of Dbp2 activity. These remodeling events are likely to be important for multiple steps of gene expression as recent studies have shown that Dbp2 couples translational efficiency of a given mRNA with the identity of the gene promoter (Espinar et al., 2018). Additionally, loss of this RNA remodeling activity may contribute to the cold-sensitive phenotype of *dbp2Δ* (Cloutier et al., 2012), due to the accumulation of stabilized, non-functional RNA structures as the temperature decreases in the absence of *DBP2*.

It is important to note that while our studies of the *YOP1* 3'UTR suggest that *DBP2*-dependent remodeling involves alteration of RNA secondary structure (Figure 2.11), remodeling includes both changes in RNA structure and RNA-binding protein composition that occurs either directly via RNA helicase activity or more indirectly by extending the RNA transcript length. It is well established that many RNA-binding proteins are inhibited by the presence of secondary structure in the targeted RNA, a phenomenon recently supported by genome-wide studies (Lambert et al., 2014). Consistently, formation of secondary structure has been shown to impair specific gene expression steps catalyzed by these RNA-binding proteins, including transcription termination (Chen and Wilusz, 1998; Klasens et al., 1998; Mortimer et al., 2014). Because loss of

DBP2 reduces recruitment of Nrd1 to individual genes and that destabilization of secondary structure in the 3'UTR of *YOP1* bypasses the requirement for *DBP2* in termination, our finding points to a model that Dbp2 remodels RNA structure to provide a “landing pad” for Nrd1. The precise mechanism for how this elicits termination and if that termination involves Nab3 and Sen1 are yet to be determined.

Although we have not detected a direct interaction between Dbp2 and Nrd1 or Nab3 to date, an interaction between Dbp2 and Sen1 has been reported along with a correlation of Dbp2 RNA-binding sites with locations of R-loop formation (Tedeschi et al., 2018). If Dbp2 enables recruitment and/or activation of the RNA-DNA helicase Sen1, this could explain prior observations that loss of *DBP2* results in accumulation of RNA-DNA hybrids *in vivo* despite reduced RNA-DNA duplex unwinding activity *in vitro* compared to pure RNA duplexes (Cloutier et al., 2016; Ma et al., 2013). While the NNS complex is not conserved in mammalian cells, Sen1 does have an orthologous protein Senataxin that functions in transcriptional termination and resolution of R-loops (Skourti-Stathaki et al., 2011). It will be interesting to determine if DDX5, the mammalian ortholog of Dbp2, interacts with Senataxin and/or the Integrator complex, the latter of which is functionally analogous to NNS (Baillat and Wagner, 2015). The fact that ectopic expression of *DDX5* in *dbp2Δ* cells rescues defects in both growth and gene expression is highly suggestive of a conserved function in both fungi and human cells (Xing et al., 2017). Moreover, mounting evidence showing that aberrant mRNA structure may underlie select human pathophysiology (Corley et al., 2015; Wan et al., 2014), points to a pressing need to understand the precise biochemical function of the ~40 DEAD-box RNA helicases in mammalian cells (~25 in budding yeast) *in vivo* (Linder and Jankowsky, 2011).

2.5 References

- Abramson, R.D., Dever, T.E., Lawson, T.G., Ray, B.K., Thach, R.E., and Merrick, W.C. (1987). The ATP-dependent interaction of eukaryotic initiation factors with mRNA. *J. Biol. Chem.* 262, 3826–3832.
- Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Séraphin, B., Hir, H. Le, and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-Box ATPase bound to RNA. *Science* (80-.). 313, 1968–1972.

- Arigo, J.T., Carroll, K.L., Ames, J.M., and Corden, J.L. (2006). Regulation of yeast NRD1 expression by premature transcription termination. *Mol. Cell* 21, 641–651.
- Arndt, K.M., and Reines, D. (2015). Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu. Rev. Biochem.* 84, 381–404.
- Aviran, S., Trapnell, C., Lucks, J.B., Mortimer, S.A., Luo, S., Schroth, G.P., Doudna, J.A., Arkin, A.P., and Pachter, L. (2011a). Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci.* 108, 11069–11074.
- Aviran, S., Lucks, J., and Pachter, L. (2011b). RNA Structure Characterization from Chemical Mapping Experiments. In *Communication, Control, and Computing (Allerton)*, 2011 49th Annual Allerton Conference on, pp. 1743–1750.
- Baejen, C., Torkler, P., Gressel, S., Essig, K., Söding, J., and Cramer, P. (2014). Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. *Mol. Cell* 55, 745–757.
- Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., et al. (2017). Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol. Cell* 66, 38–49.e6.
- Baillat, D., and Wagner, E.J. (2015). Integrator: Surprisingly diverse functions in gene expression. *Trends Biochem. Sci.* 40, 257–264.
- Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Séraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* 12, 861–869.
- Barta, I., and Iggo, R. (1995). Autoregulation of expression of the yeast Dbp2p “DEAD-box” protein is mediated by sequences in the conserved DBP2 intron. *EMBO J.* 14, 3800–3808.
- Beck, Z.T., Cloutier, S.C., Schipma, M.J., Petell, C.J., Ma, W.K., and Tran, E.J. (2014). Regulation of glucose-dependent gene expression by the RNA helicase Dbp2 in *Saccharomyces cerevisiae*. *Genetics* 198, 1001–1014.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 426–438.

- Chen, F., and Wilusz, J. (1998). Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. *Nucleic Acids Res.* *26*, 2891–2898.
- Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in *Saccharomyces cerevisiae*. *Eukaryot. Cell* *11*, 417–429.
- Choudhary, K., Shih, N.P., Deng, F., Ledda, M., Li, B., and Aviran, S. (2016). Metrics for rapid quality control in RNA structure probing experiments. *Bioinformatics* *32*, 3575–3583.
- Choudhary, K., Lai, Y.-H., Tran, E.J., and Aviran, S. (2019). dStruct: identifying differentially reactive regions from RNA structurome profiling data. *Genome Biol.* *20*, 40.
- Cloutier, S.C., Ma, W.K., Nguyen, L.T., and Tran, E.J. (2012). The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. *J. Biol. Chem.* *287*, 26155–26166.
- Cloutier, S.C., Wang, S., Ma, W.K., Petell, C.J., and Tran, E.J. (2013). Long Noncoding RNAs Promote Transcriptional Poising of Inducible Genes. *PLoS Biol.* *11*, 32–34.
- Cloutier, S.C., Wang, S., Ma, W.K., Al Husini, N., Dhoondia, Z., Ansari, A., Pascuzzi, P.E., and Tran, E.J. (2016). Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. *Mol. Cell* *61*, 393–404.
- Conti, E., Han, Z., Porrua, O., Leonaitė, B., Basquin, J., Bonneau, F., and Libri, D. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *EMBO J.* *36*, 1590–1604.
- Corley, M., Solem, A., Qu, K., Chang, H.Y., and Laederach, A. (2015). Detecting riboSNitches with RNA folding algorithms: A genome-wide benchmark. *Nucleic Acids Res.* *43*, 1859–1868.
- Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.* *7*, e1002329.
- Dardenne, E., PolayEspinoza, M., Fattet, L., Germann, S., Lambert, M.P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., et al. (2014). RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Rep.* *7*, 1900–1913.

- Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700.
- Ding, Y., Kwok, C.K., Tang, Y., Bevilacqua, P.C., and Assmann, S.M. (2015). Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* 10, 1050–1066.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dunn, E.F., Hammell, C.M., Hodge, C.A., and Cole, C.N. (2005). Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. *Genes Dev.* 19, 90–103.
- Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* 113, 8456–8490.
- Espinar, L., Tamarit, M.À.S., Domingo, J., and Carey, L.B. (2018). Promoter architecture determines cotranslational regulation of mRNA. *Genome Res.* 28, 509–518.
- Fairman, M.E., Maroney, P.A., Wang, W., Bowers, H.A., Gollnick, P., Nilsen, T.W., and Jankowsky, E. (2004). Protein displacement by DExH/D “RNA helicases” without duplex unwinding. *Science* 304, 730–734.
- Flynn, R.A., Martin, L., Spitale, R.C., Do, B.T., Sagan, S.M., Zarnegar, B., Qu, K., Khavari, P.A., Quake, S.R., Sarnow, P., et al. (2015). Dissecting noncoding and pathogen RNA-protein interactomes. *RNA* 21, 135–143.
- Flynn, R.A., Zhang, Q.C., Spitale, R.C., Lee, B., Mumbach, M.R., and Chang, H.Y. (2016). Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.* 11, 273–290.
- Gelbart, M.E., Rechsteiner, T., Richmond, T.J., and Tsukiyama, T. (2001). Interactions of Isw2 chromatin remodeling complex with nucleosomal arrays: analyses using recombinant yeast histones and immobilized templates. *Mol. Cell. Biol.* 21, 2098–2106.
- Gilman, B., Tijerina, P., and Russell, R. (2017). Distinct RNA-unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs. *Biochem. Soc. Trans.* 45, 1313–1321.

- Grzechnik, P., and Kufel, J. (2008). Polyadenylation Linked to Transcription Termination Directs the Processing of snoRNA Precursors in Yeast. *Mol. Cell* 32, 247–258.
- Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev.* 29, 849–861.
- Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. *Mol. Cell* 48, 409–421.
- Guenther, U.P., Weinberg, D.E., Zubradt, M.M., Tedeschi, F.A., Stawicki, B.N., Zagore, L.L., Brar, G.A., Licatalosi, D.D., Bartel, D.P., Weissman, J.S., et al. (2018). The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature* 559, 130–134.
- Hammell, C.M., Gross, S., Zenklusen, D., Heath, V., Stutz, F., Moore, C., Cole, C.N., and Heath, C. V (2002). Coupling of Termination , 3' Processing , and mRNA Export. *Mol. Cell. Biol.* 22, 6441–6457.
- Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res.* 45, 1355–1370.
- Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* 18, 263–273.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* 38, 576–589.
- Hirling, H., Scheffner, M., Restle, T., and Stahl, H. (1989). RNA helicase activity associated with the human p68 protein. *Nature* 339, 562–564.
- Jamonnak, N., Creamer, T.J., Darby, M.M., Schaughency, P., Wheelan, S.J., and Corden, J.L. (2011). Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* 17, 2011–2025.
- Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S., and Wu, J.Y. (2011). RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol. Cell. Biol.* 31, 1812–1821.

- Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A., and Moore, C.L. (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 Dev.* *11*, 2545–2556.
- Kim, K.Y., and Levin, D.E. (2011). Mpk1 MAPK association with the paf1 complex blocks sen1-mediated premature transcription termination. *Cell* *144*, 745–756.
- Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat. Struct. Mol. Biol.* *17*, 1279–1286.
- Klasens, B.I.F., Das, A.T., and Berkhout, B. (1998). Inhibition of polyadenylation by stable RNA secondary structure. *Nucleic Acids Res.* *26*, 1870–1876.
- König, J., Zarnack, K., Rot, G., Curk, T.T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* *17*, 909–915.
- Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* *26*, 1891–1896.
- Kuehner, J.N., and Brow, D.A. (2008). Regulation of a Eukaryotic Gene by GTP-Dependent Start Site Selection and Transcription Attenuation. *Mol. Cell* *31*, 201–211.
- Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. *Mol. Cell* *54*, 887–900.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.
- Larochelle, M., Robert, M.A., Hébert, J.N., Liu, X., Matteau, D., Rodrigue, S., Tian, B., Jacques, P.É., and Bachand, F. (2018). Common mechanism of transcription termination at coding and noncoding RNA genes in fission yeast. *Nat. Commun.* *9*, 4364.
- Lemay, J.F., Marguerat, S., Larochelle, M., Liu, X., van Nues, R., Hunyadkúrti, J., Hoque, M., Tian, B., Granneman, S., Bähler, J., et al. (2016). The Nrd1-like protein Seb1 coordinates cotranscriptional 3' end processing and polyadenylation site selection. *Genes Dev.* *30*, 1558–1572.

- Lesser, C.F., and Guthrie, C. (1993). Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1. *Genetics* 133, 851–863.
- Liebeg, A., Mayer, O., and Waldsich, C. (2010). DEAD-box protein facilitated RNA folding in vivo. *Rna Biol.* 7, 803–811.
- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* 12, 505–516.
- Liu, X., Hoque, M., Larochelle, M., Lemay, J.F., Yurko, N., Manley, J.L., Bachand, F., and Tian, B. (2017). Comparative analysis of alternative polyadenylation in *S. Cerevisiae* and *S. Pombe*. *Genome Res.* 27, 1685–1695.
- Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26.
- Ma, W.K., Cloutier, S.C., and Tran, E.J. (2013). The DEAD-box protein Dbp2 functions with the RNA-binding protein Yra1 to promote mRNP assembly. *J. Mol. Biol.* 425, 3824–3838.
- Ma, W.K., Paudel, B.P., Xing, Z., Sabath, I.G., Rueda, D., and Tran, E.J. (2016). Recruitment, Duplex Unwinding and Protein-Mediated Inhibition of the Dead-Box RNA Helicase Dbp2 at Actively Transcribed Chromatin. *J. Mol. Biol.* 428, 1091–1106.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10.
- Martin, R., Straub, A.U., Doebele, C., and Bohnsack, M.T. (2013). DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol.* 10, 4–18.
- Mischo, H.E., and Proudfoot, N.J. (2013). Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1829, 174–185.
- Moqtaderi, Z., Geisberg, J. V., and Struhl, K. (2018). Extensive Structural Differences of Closely Related 3' mRNA Isoforms: Links to Pab1 Binding and mRNA Stability. *Mol. Cell* 72, 849–861.e6.
- Mortimer, S. a, Kidwell, M.A., and Doudna, J. a (2014). Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* 15, 469–479.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* (80-.). 320, 1344–1349.

- Nielsen, K.H., Chamieh, H., Andersen, C.B.F., Fredslund, F., Hamborg, K., Le Hir, H., and Andersen, G.R. (2008). Mechanism of ATP turnover inhibition in the EJC. *RNA* *15*, 67–75.
- Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A., and Ramos, A. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat. Struct. Mol. Biol.* *12*, 144–151.
- Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., and Milos, P.M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* *143*, 1018–1029.
- Parsa, J.Y., Boudoukha, S., Burke, J., Homer, C., and Madhani, H.D. (2018). Polymerase pausing induced by sequence-specific RNA-binding protein drives heterochromatin assembly. *Genes Dev.* *32*, 953–964.
- Peattie, D.A., and Gilbert, W. (1980). Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. U. S. A.* *77*, 4679–4682.
- Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* *497*, 127–131.
- Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: Why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* *16*, 190–202.
- Potratz, J.P., Del Campo, M., Wolf, R.Z., Lambowitz, A.M., and Russell, R. (2011). ATP-dependent roles of the DEAD-box protein Mss116p in group II intron splicing in vitro and in vivo. *J. Mol. Biol.* *411*, 661–679.
- Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* (80-.). *352*, aad9926-aad9926.
- Putnam, A.A., and Jankowsky, E. (2013). DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim. Biophys. Acta - Gene Regul. Mech.* *1829*, 884–893.
- Qu, X., Lykke-Andersen, S., Nasser, T., Saguez, C., Bertrand, E., Jensen, T.H., and Moore, C. (2009). Assembly of an export-competent mRNP is needed for efficient release of the 3'-end processing complex after polyadenylation. *Mol. Cell. Biol.* *29*, 5327–5338.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* *44*, W160–W165.

- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Rogers, G.W., Richter, N.J., and Merrick, W.C. (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem.* 274, 12236–12244.
- Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-Safe Transcriptional Termination for Protein-Coding Genes in *S. cerevisiae*. *Mol. Cell* 36, 88–98.
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2013). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705.
- Rudolph, M.G., and Klostermeier, D. (2015). When core competence is not enough: Functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. *Biol. Chem.* 396, 849–865.
- Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-Wide Mapping of Yeast RNA Polymerase II Termination. *PLoS Genet.* 10, e1004632.
- Singh, N.N., Singh, R.N., and Androphy, E.J. (2007). Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res.* 35, 371–389.
- Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol. Cell* 42, 794–805.
- Sloma, M.F., and Mathews, D.H. (2015). Improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol.* 553, 91–114.
- Steinmetz, E.J., and Brow, D.A. (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 Brow, D.A. (2003). Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol* 23, 6339–6349.

- Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413, 327–331.
- Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase. *Mol. Cell* 24, 735–746.
- Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* 13, R67.
- Suh, H., Ficarro, S.B., Kang, U.-B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Mol. Cell* 61, 297–304.
- Tedeschi, F.A., Cloutier, S.C., Tran, E.J., and Jankowsky, E. (2018). The DEAD-box protein Dbp2p is linked to noncoding RNAs, the helicase Sen1p, and R-loops. *RNA* 24, 1693–1705.
- Tijerina, P., Bhaskaran, H., and Russell, R. (2006). Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. *Proc. Natl. Acad. Sci. U. S. A.* 103, 16698–16703.
- Tran, E.J., Zhou, Y., Corbett, A.H., and Wentz, S.R. (2007). The DEAD-Box Protein Dbp5 Controls mRNA Export by Triggering Specific RNA:Protein Remodeling Events. *Mol. Cell* 28, 850–859.
- Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* 15, 795–804.
- Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709.
- Wang, S., Xing, Z., Pascuzzi, P.E., and Tran, E.J. (2017). Metabolic Adaptation to Nutrients Involves Coregulation of Gene Expression by the RNA Helicase Dbp2 and the Cyc8 Corepressor in *Saccharomyces cerevisiae*. *G3 Genes, Genomes, Genet.* 7, 2235–2247.

- Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* *15*, R8.
- Wilmes, G.M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., Collins, S.R., Whitworth, G.B., Kress, T.L., Weissman, J.S., et al. (2008). A Genetic Interaction Map of RNA-Processing Factors Reveals Links between Sem1/Dss1-Containing Complexes and mRNA Export and Splicing. *Mol. Cell* *32*, 735–746.
- Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* *30*, 1790–1803.
- Xing, Z., Wang, S., and Tran, E.J. (2017). Characterization of the mammalian DEAD-box protein DDX5 reveals functional conservation with *S. cerevisiae* ortholog Dbp2 in transcriptional control and glucose metabolism. *RNA* *23*, 1125–1138.
- Xing, Z., Ma, W.K., and Tran, E.J. (2018). The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscip. Rev. RNA* *0*, e1519.
- Yang, Q., and Jankowsky, E. (2005). ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* *44*, 13591–13601.
- Yang, Q., and Jankowsky, E. (2006). The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Mol. Biol.* *13*, 981–986.
- Yassour, M., Kaplan, T., Fraser, H.B., Levin, J.Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtkova, I., Gnirke, A., et al. (2009). Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proc. Natl. Acad. Sci.* *106*, 3264–3269.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* *9*, R137.

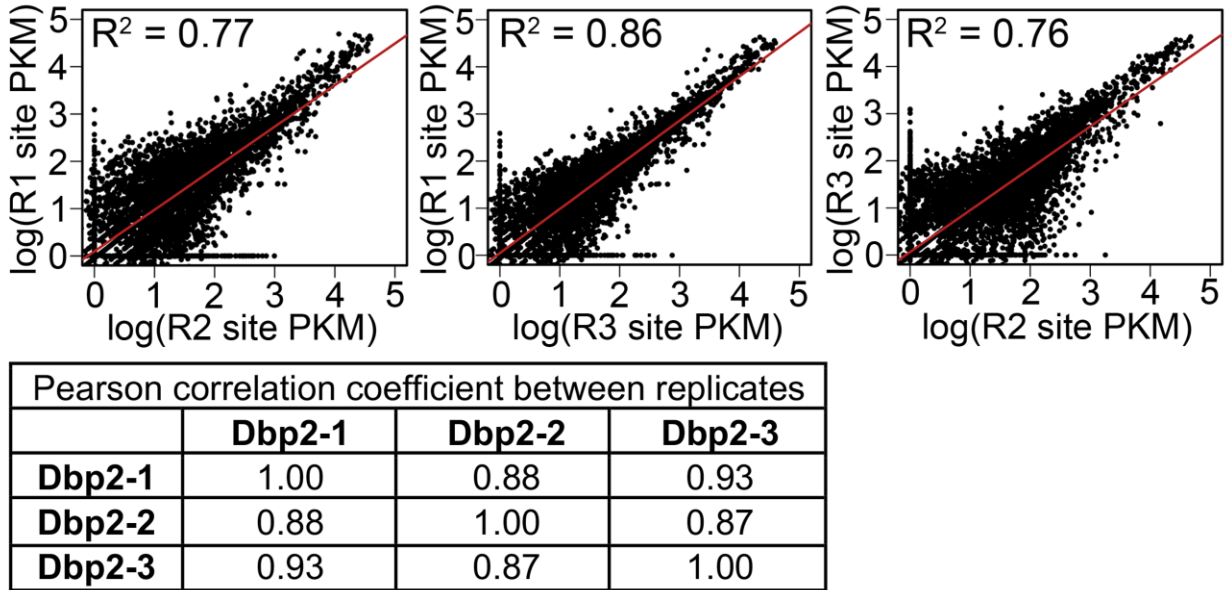
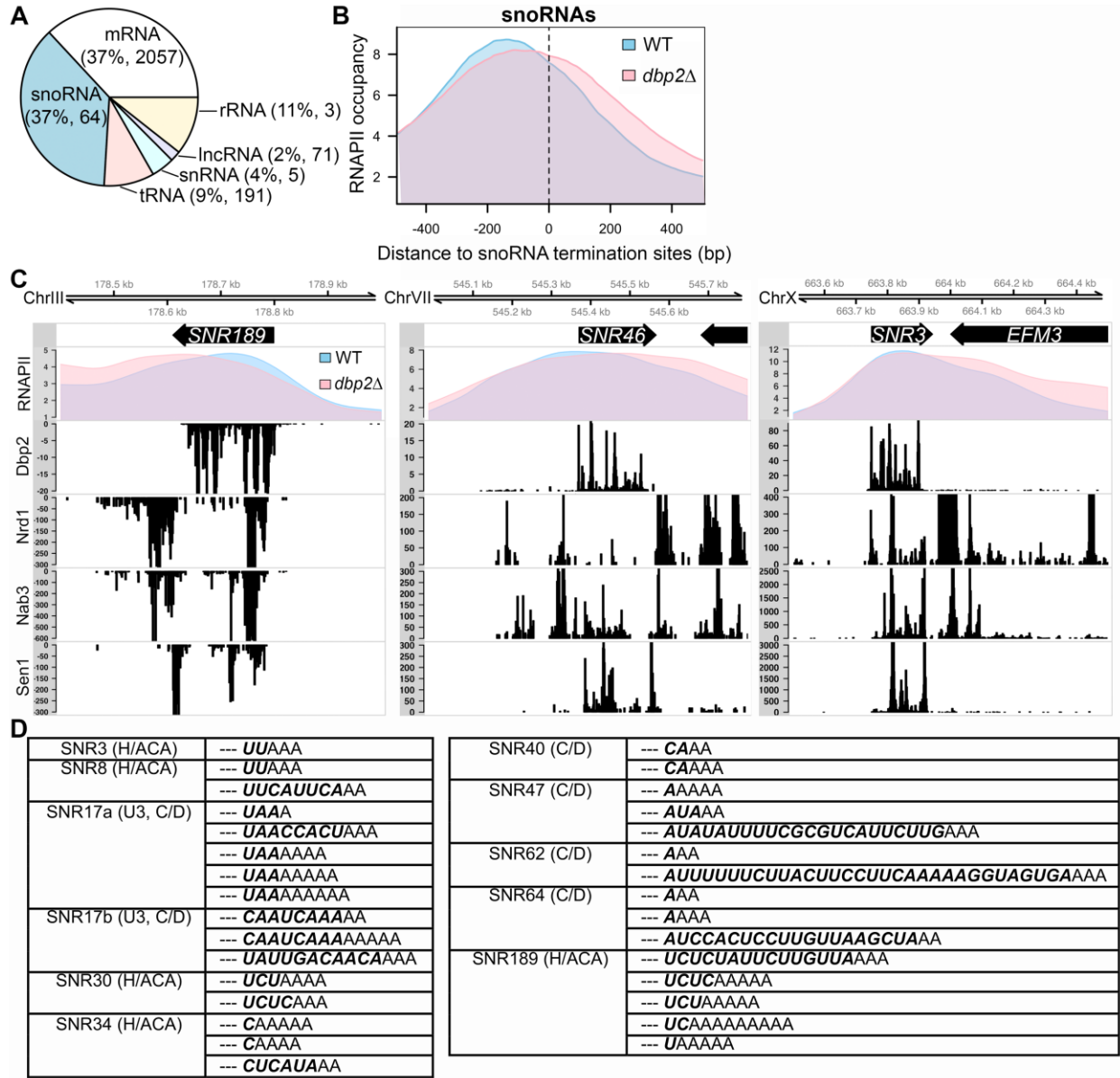


Figure 2.1. Assessment of reproducibility of Dbp2-binding sites as determined by iCLIP-seq across three biological replicates (R1, R2, and R3).

For each replicate, the crosslinking count per kilobase of transcript per million mapped reads (PKM) was calculated. Reproducibility was assessed by calculating Pearson's correlation coefficient between each pair of replicates (table below).

Figure 2.2. Dbp2 promotes transcription termination and processing at snoRNA genes.

(A) *Dbp2 binds both coding and non-coding RNAPII transcripts.* The composition of RNA classes identified by native iCLIP-seq of strains harboring C-terminally 3XFLAG-tagged *DBP2* at the endogenous locus. Dbp2-binding sites across all RNA classes were derived from uniquely mapped reads in each replicate. The percentage of the total Dbp2-bound RNAs and number of unique transcripts in each class (shown in parentheses) were calculated and averaged from the three biological replicates. Note that the number of rRNA transcripts in Dbp2 iCLIP-seq is under-represented due to rDNA repeats. (B) *RNAPII ChIP-seq reveals snoRNA gene termination defects in *dbp2Δ* cells.* Normalized RNAPII occupancy across termination sites of monocistronic snoRNA genes from RNAPII ChIP-seq in the wild type (blue) and *dbp2Δ* (pink). SnoRNA termination sites were obtained from (Schaughency et al., 2014). (C) *Comparison of RNAPII occupancy with RNA-binding sites of Dbp2 and Nrd1, Nab3, and Sen1 at individual snoRNA gene loci.* Normalized RNAPII profiles across select snoRNA genes from RNAPII ChIP-seq in wild type and *dbp2Δ* (top panel). The binding patterns of Nrd1, Nab3, and Sen1 were reproduced from previously published data (Creamer et al., 2011). Note that the difference in scales is due to the difference in how data was generated and does not represent the absolute quantity of binding. (D) *Unprocessed snoRNA sequences in Dbp2 iCLIP-seq reads.* Table of all Dbp2-bound snoRNA sequences that show unprocessed nucleotides (bold italics) and non-templated A's at 3' ends of the mature snoRNA (dashed lines). The class of snoRNA (Box C/D or Box H/ACA) are listed in parentheses next to the name.



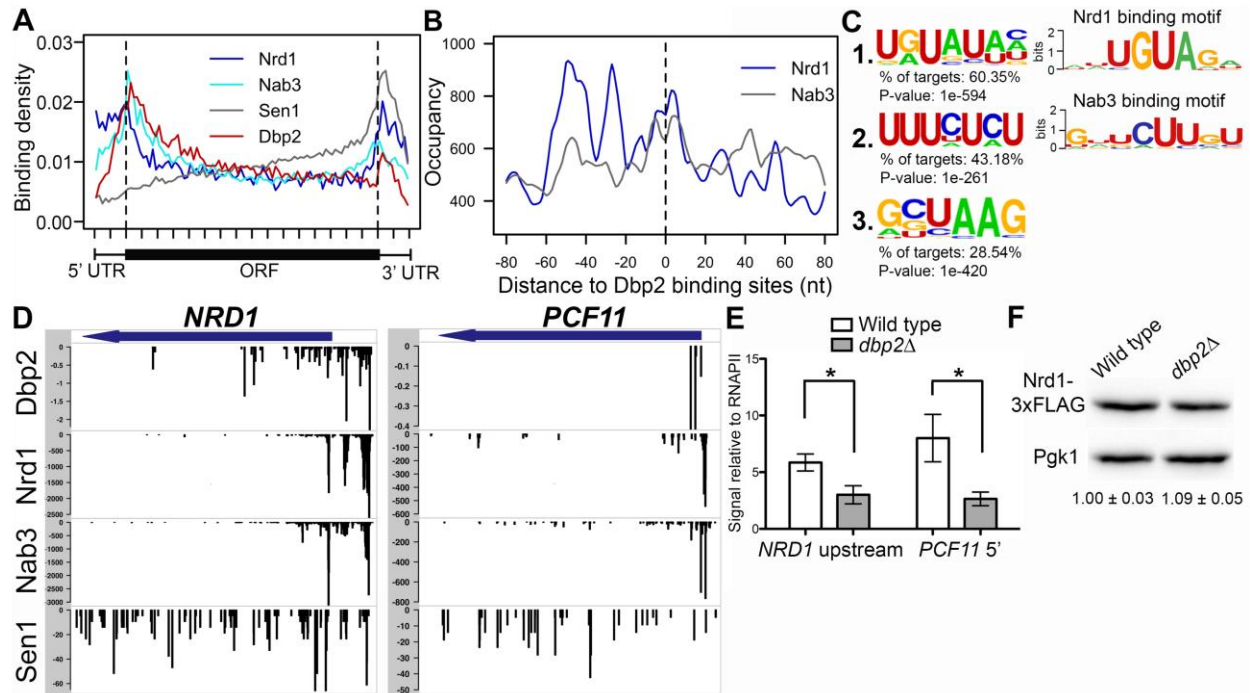


Figure 2.3. Dbp2 binding in protein-coding transcripts correlates with Nrd1 and Nab3 binding sites and shares similar RNA sequence motifs.

(A) *Dbp2*, *Nrd1*, *Nab3*, and *Sen1* meta-analysis reveals similar distribution patterns on protein-coding genes. Meta-analysis of *Nrd1* (blue), *Nab3* (cyan), *Sen1* (gray) (Creamer et al., 2011), and *Dbp2* (red) binding sites across all commonly bound mRNAs. Dashed vertical lines mark boundaries between 5' or 3' UTRs and ORFs. (B) *Nrd1* and *Nab3* show enriched occupancy at *Dbp2*-binding sites across protein-coding transcripts. The distance between *Dbp2* and *Nrd1* (blue) or *Nab3* (gray) binding sites in all commonly bound mRNAs. (C) Enriched sequence motifs bound by *Dbp2* in protein-coding transcripts and comparison to known *Nrd1* and *Nab3*-binding sites. Motifs were identified from *Dbp2* iCLIP-seq using HOMER (Heinz et al., 2010). Web logos of sequence motifs bound by *Nrd1* and *Nab3* were reproduced from published reports (Creamer et al., 2011) (right). (D) Comparison of *Dbp2* binding with *Nrd1*, *Nab3*, and *Sen1* binding at previously reported protein-coding targets of the NNS complex. Meta-analysis of *Nrd1*, *Nab3*, and *Sen1* RNA-binding data sets from Creamer et al in conjunction with *Dbp2*. (Creamer et al., 2011). (E) Loss of *DBP2* reduces recruitment of *Nrd1* to protein-coding genes. ChIP-qPCR of endogenously, 3XFLAG-tagged *NRD1* strain at 5' ends of *NRD1* (left) and *PCF11* (right). Signals are shown relative to input and RNAPII ChIP levels, the latter of which accounts for changes in abundance due to transcriptional activity. Asterisks (*) indicate two-sided p-value < 0.05. (F) Reduced *Nrd1* association is not a result of decreased *Nrd1*-3XFLAG levels in *dbp2Δ*. Representative western blot of *Nrd1*-3XFLAG and *Pgk1* in wild type and *dbp2Δ* strains. Numbers below correspond to the relative level of *Nrd1*-3XFLAG in *dbp2Δ* cells compared to wild type normalized to a *Pgk1* loading control across three independent biological replicates. The signal of the wild type is set to one and standard deviations are shown.

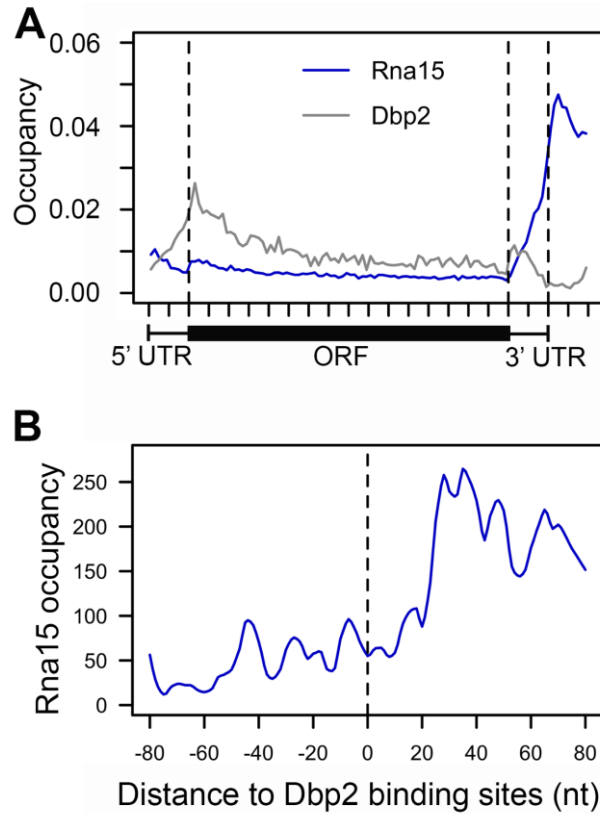


Figure 2.4. Dbp2 binding at the 3' ends of protein-coding transcripts does not correlate with Rna15 component of the cleavage and polyadenylation complex (CPF).

(A) Meta-analysis of Dbp2 (grey) and Rna15 (blue) RNA-binding sites across all commonly bound mRNAs (Rna15 sites from (Baejen et al., 2014)). (B) The distance between Dbp2 and Rna15 binding sites in all commonly bound mRNAs.

Figure 2.5. Loss of *DBP2* causes termination defects at a subset of protein-coding genes.

(A) *Meta-analysis of normalized RNAPII occupancy at the 3' ends of all protein-coding genes in wild type and *dbp2Δ* cells.* Only transcripts without overlapping genes within 150 bp downstream in the sense direction were considered in the analysis. The 0 position and dotted line marks the location of the polyadenylation site (Nagalakshmi et al., 2008; Yassour et al., 2009).

(B) *Venn diagram showing the intersection between 3' extended transcripts in *dbp2Δ* and mRNAs bound by Dbp2 at the 3' end (50 nt of ORF 3' end through the 3' UTR).* Aberrant transcripts were identified following RNA-seq of wild-type and *dbp2Δ* strains by an overaccumulation of reads mapping at least 150 nts downstream of the 3' UTR, after accounting for different expression levels between wild type and *dbp2Δ* strains. The number in the parentheses is the expected value of intersection if the two groups of transcripts have no significant correlation. The number within the white square corresponds to genes that lack both detectable, putative read-through products in *dbp2Δ* and Dbp2-binding on mRNA 3' ends. The p-value derived from a one-sided Fisher's exact test is shown.

(C) *Loss of *DBP2* results in differential accumulation of 3' extended products for individual genes.* Histograms illustrating the percentage of total transcripts corresponding to 3' extended products in *dbp2Δ* or wild type cells as determined by read counts in the extended region over the counts in the ORF in wild type and *dbp2Δ* cells multiplied by 100. Grey coloring denotes overlap between the histogram of wild type and *dbp2Δ*.

(D) *Loss of *DBP2* results in a broader distribution of 3' extended transcripts as compared to wild type cells.* A box plot showing the quartile distribution of the ratio of extended versus total transcript in wild type and *dbp2Δ* cells. The distributions of the two strains is significantly different, as tested by the two sample Kolmogorov–Smirnov test (p-value < 2.2e-16).

(E) *Protein-coding genes that produce 3' extended transcripts show RNAPII accumulation in *dbp2Δ* cells at downstream of annotated 3' UTRs.* Meta-analysis of RNAPII occupancy across the 824 genes with 3' extended transcripts in *dbp2Δ* RNA seq.

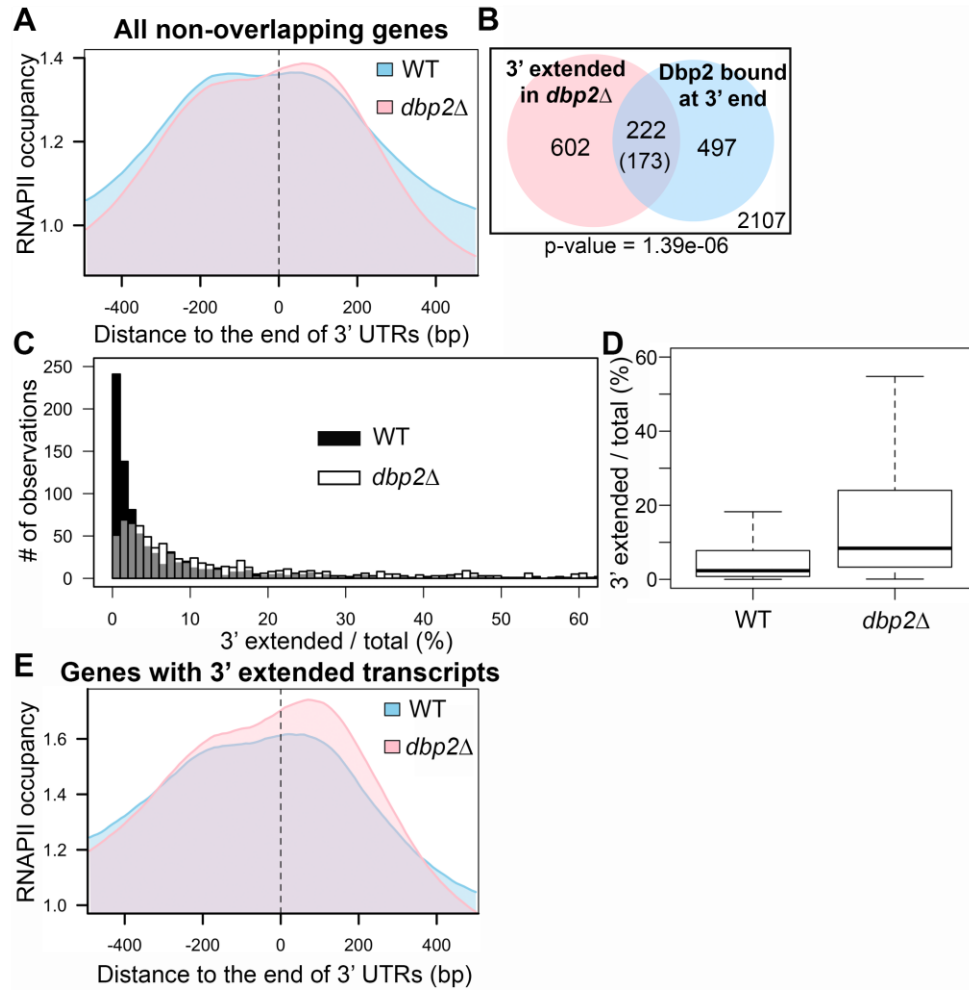
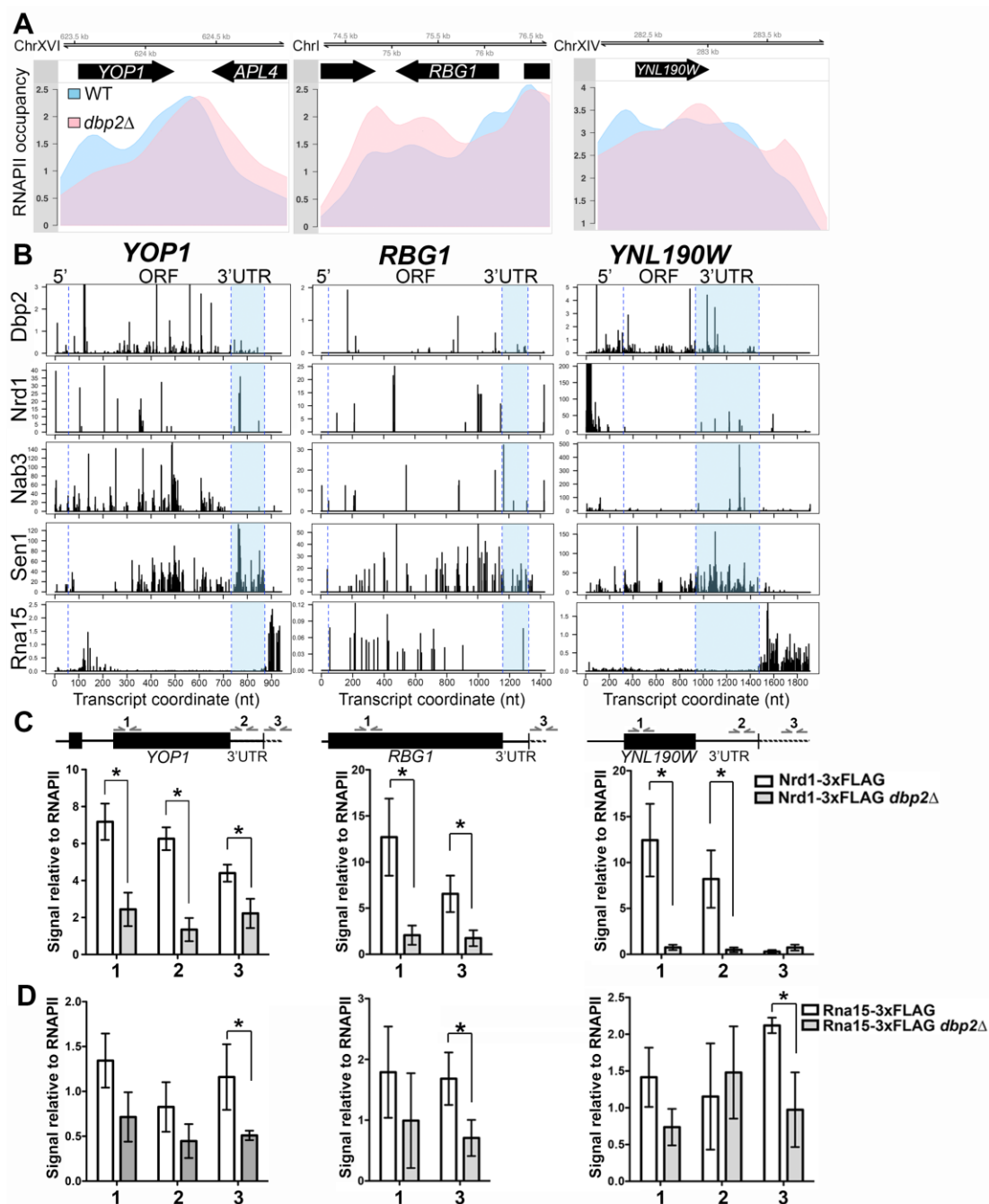


Figure 2.6. DBP2-dependent termination of protein-coding genes correlates with efficient DBP2-dependent recruitment of Nrd1 within the gene ORF and 3'UTR.

(A) *RNAPII occupancy at three, selected protein-coding genes shows DBP2-dependent termination.* Normalized RNAPII occupancy at the three termination-defective protein-coding genes in the wild type (blue) and *dbp2Δ* (pink). (B) *Comparison of Dbp2, Nrd1, Nab3, Sen1, and Rna15 RNA-binding sites at individual genes shows similar distributions between Dbp2 and members of the NNS complex.* The published binding sites for Nrd1, Nab3, Sen1 and Rna15 were obtained from the GEO session GSE31764 (Creamer et al., 2011) and GSE59676 (Baejen et al., 2014). The blue dashed lines correspond to the boundaries of the coding region and the end of 3' UTR and are aligned to the schematic representation in (C). (C) *Loss of DBP2 reduces recruitment of Nrd1 to protein-coding genes.* ChIP-qPCR of endogenously, 3XFLAG-tagged *NRD1* strain at 5' ends (1), 3'UTRs (2), and downstream of polyadenylation sites (3) in *YOP1*, *RBG1*, and *YNL190W*. Signals are shown relative to input and RNAPII ChIP levels, the latter of which accounts for changes in abundance due to transcriptional activity. Asterisks (*) indicate two-sided p-value < 0.05. Note that the 3'UTR (2) of *RBG1* was not assayed due to a technical limitation of qPCR primer binding sites. (D) *Rna15 shows reduced chromatin binding in dbp2Δ cells but only downstream of the polyadenylation site.* ChIP-qPCR was conducted as in C but with an endogenously 3XFLAG-tagged *RNA15* strain.



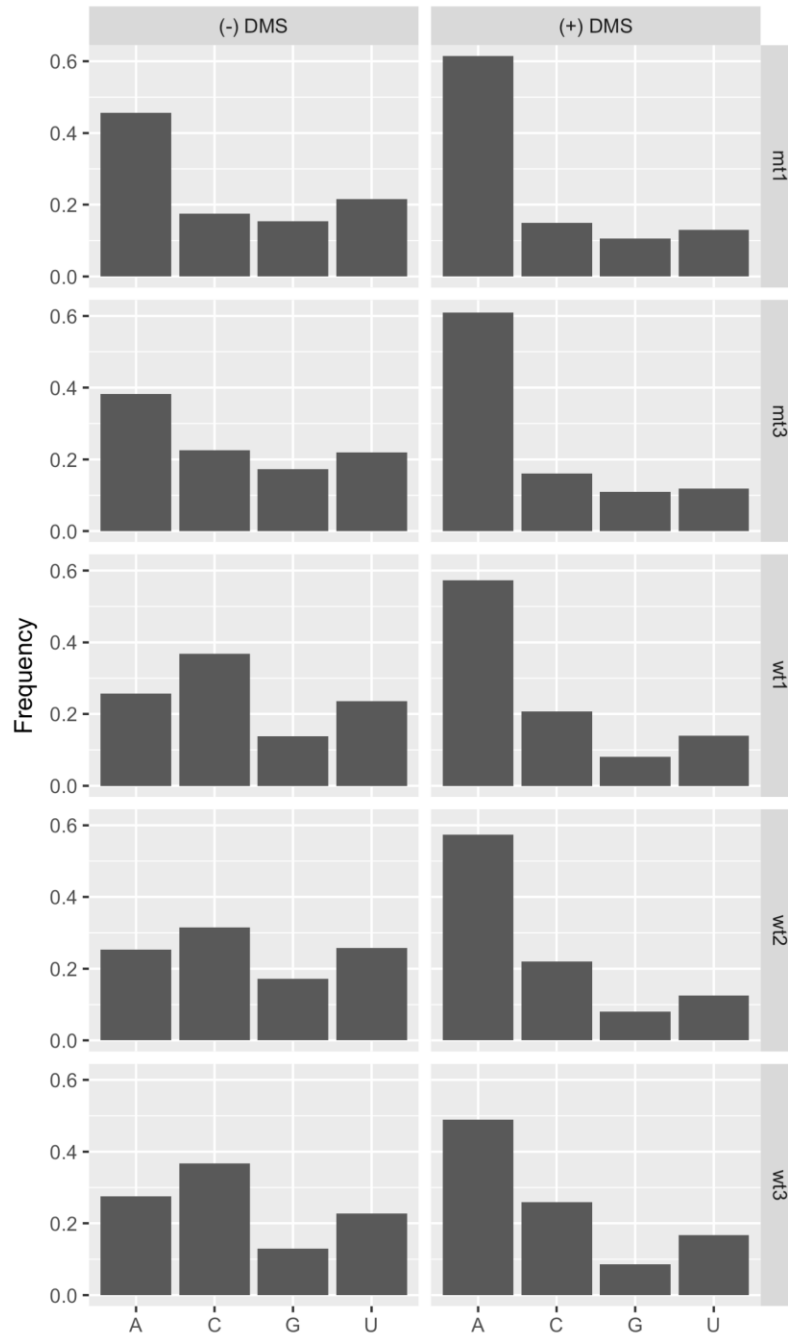
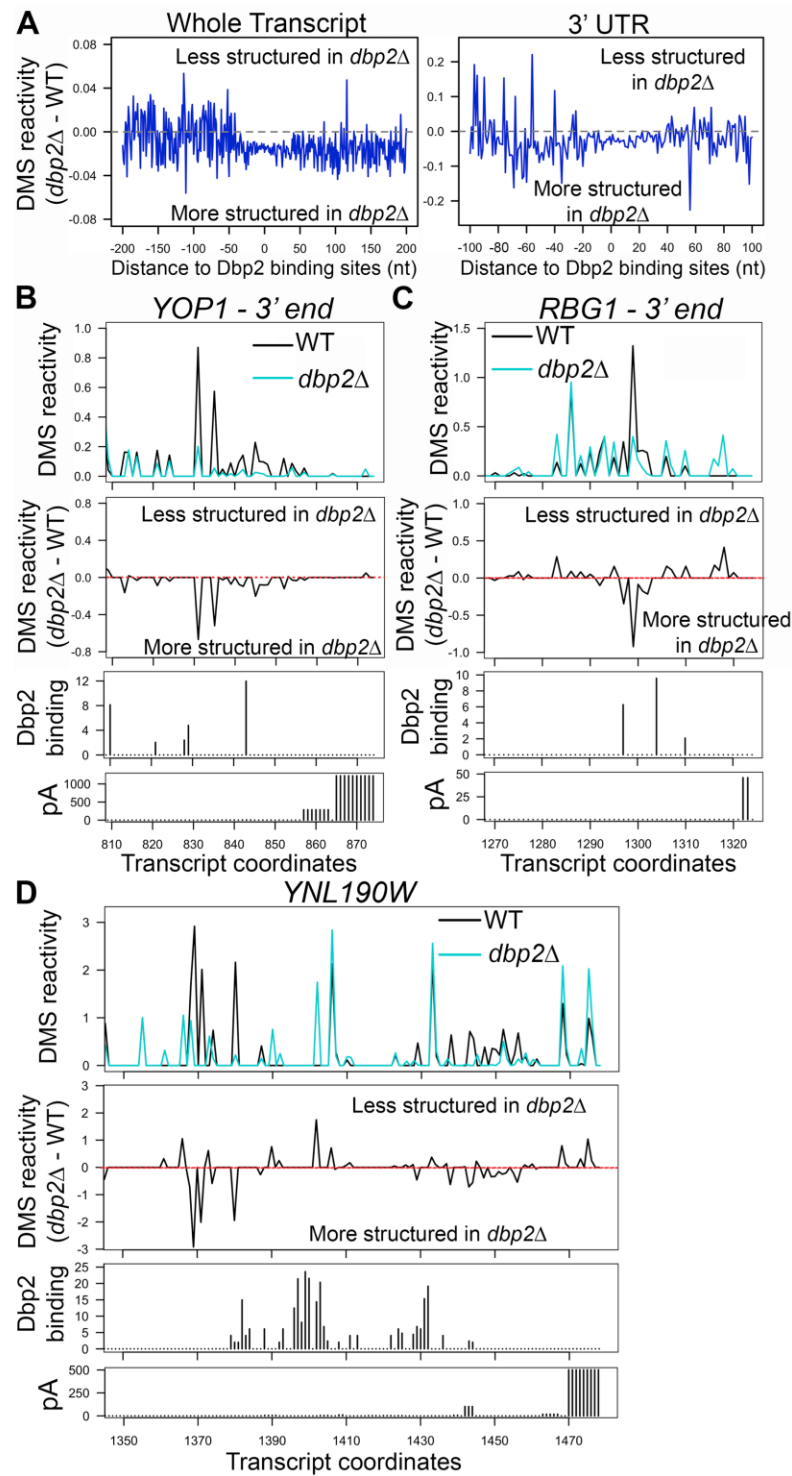


Figure 2.7. The frequency of RT stops at A, U, C, and G in each replicate of DMS-treated (+) and untreated (-) samples.

Vertical strip labels indicate strain ('mt' for *dbp2Δ* and 'wt' for wild type) and replicate batch number. Note reactivity bias towards A, consistent with prior studies (Ding et al., 2014; Rouskin et al., 2013).

Figure 2.8. Structure-seq reveals *DBP2*-dependent RNA structural changes in protein-coding genes.

(A) Meta-analyses of DMS reactivity at Dbp2-binding sites on mRNAs reveals 40-60 nt “trough” of *DBP2*-dependent structure. The relationship between Dbp2-binding sites and the changes in DMS reactivity was analyzed for both the whole transcript (left panel) and 3' UTR (right panel). *DBP2*-dependent change in DMS reactivity of each nucleotide in Dbp2-bound mRNA transcripts was calculated by subtracting the average DMS reactivity in *dbp2Δ* by that in wild type, and was then plotted corresponding to the distance to Dbp2-binding sites identified in iCLIP-seq. (B-D) Reactivity profiles of the 3' ends of *YOP1* (B), *RBG1* (C), and *YNL190W* (D) in wild type and *dbp2Δ* cells (top panels). The x-axes indicate transcript coordinates relative to the transcriptional start site. The annotated 3' UTR region spans coordinates 733-872 for *YOP1*, 1156-1322 for *RBG1*, 934-1476 for *YNL190W* were defined previously (Nagalakshmi et al., 2008; Yassour et al., 2009). Only the regions close to the end of 3' UTR are shown (50 – 120 nt). *DBP2*-dependent changes in DMS reactivity are also presented as the values derived by subtracting the reactivity in wild type from the reactivity in *dbp2Δ* (middle panels). Dbp2 binding sites derived from iCLIP-seq are shown below reactivity profiles. Annotated polyadenylation sites (pA, (Ozsolak et al., 2010)) are shown for reference (bottom panels). Bar height in the pA histograms indicate the relative frequency of polyadenylation at each position.



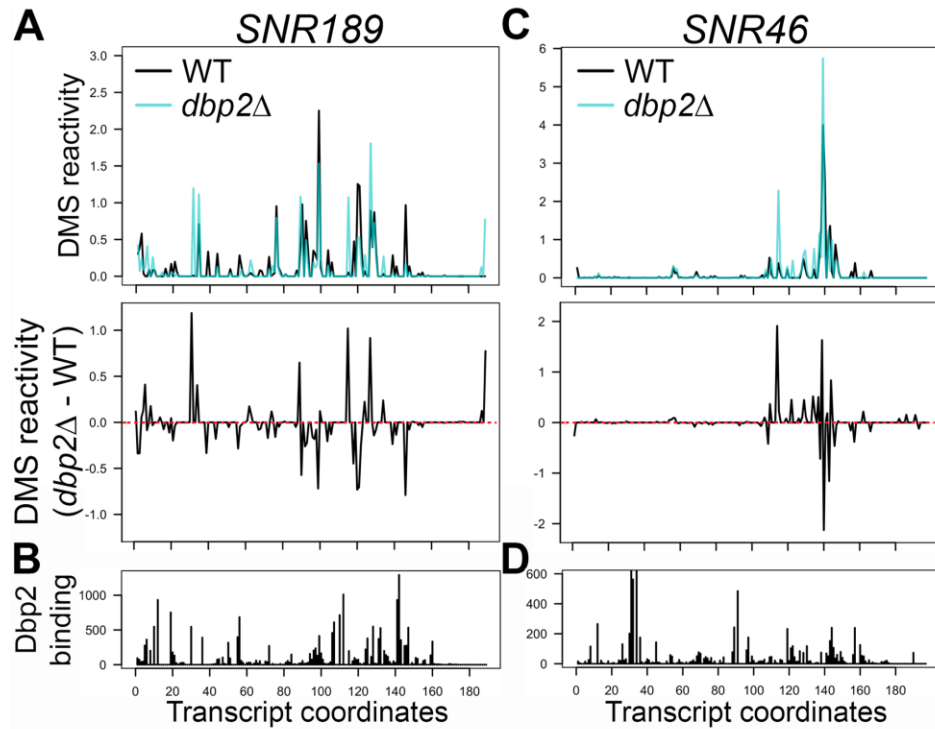


Figure 2.9. DMS reactivity (A, C) and Dbp2 binding (B, D) profiles of the two snoRNAs with a read-through defect in *dbp2Δ*.

The reactivity towards DMS for each nucleotide was normalized using 2-8% approach in each replicate, and the value shown on the Y-axis is the average from biological replicates of the same strain (i.e. wild type or *dbp2Δ*). The numbers at the X-axis represent the transcript coordinates relative to the start of mature 5' end, and the end position on the X-axis corresponds to the mature 3' end. *DBP2*-dependent changes in DMS reactivity are also presented as the values derived by subtracting the reactivity in wild type from the reactivity in *dbp2Δ* (A, bottom). Note that SNR189 also exhibits Dbp2-dependent processing as evidenced by non-templated A's in *dbp2Δ* (Figure 1E).

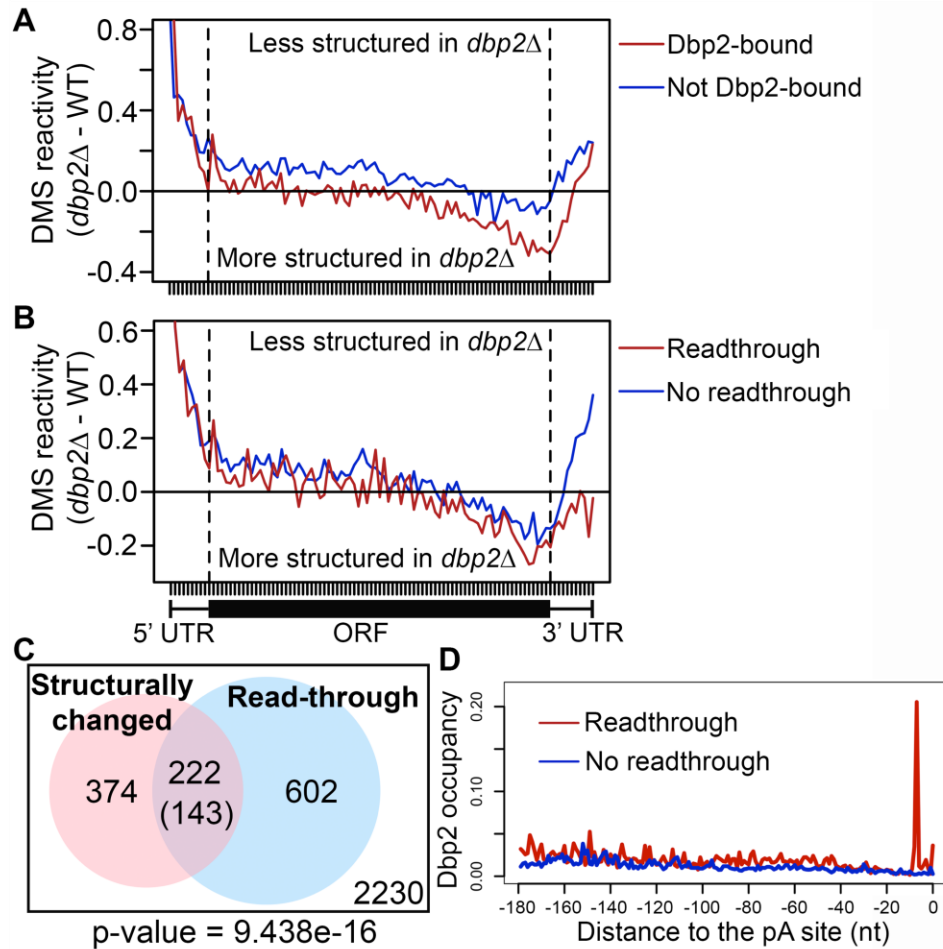


Figure 2.10. The presence of *DBP2*-dependent structural changes in 3' UTRs correlates with a requirement for *DBP2* in transcriptional termination.

(A) Metagene analysis of *DBP2*-dependent changes on secondary structure in transcripts bound (red) or not bound (blue) by Dbp2. *DBP2*-dependent structural changes were captured by the differential reactivity of DMS in *dbp2Δ* versus wild type cells. (B) Metagene analysis of *DBP2*-dependent DMS reactivity changes as plotted with respect to the presence (red) or absence (blue) of a read-through defect in *dbp2Δ*. (C) A Venn diagram showing the intersection between transcripts with read-through defects and those with significant structural changes in *dbp2Δ* (determined using dStruct described in methods). The number in the parentheses is the expected value of intersection if the two groups of transcripts have no significant relationship. The p-value derived from a one-sided Fisher's exact test is shown below the box. The number shown in the square corresponds to transcripts without a read-through defect and a statistically significant, structural change in *dbp2Δ*. In this analysis, only transcripts that were passed to the read-through and DMS reactivity analyses are included. (D) Meta-analysis of Dbp2 binding within 200 nt upstream of annotated polyadenylation sites of mRNAs with (red) or without 3' extension (blue) upon *DBP2* deletion.

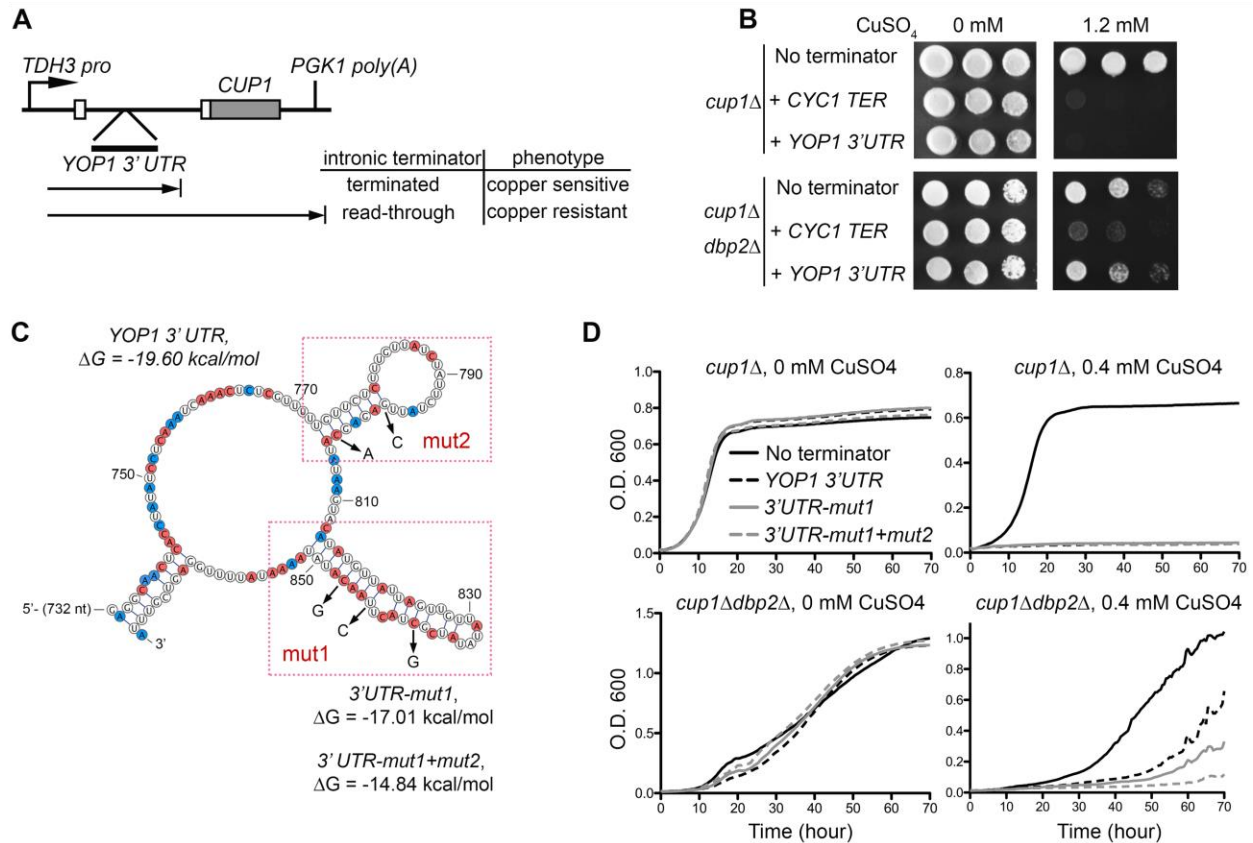


Figure 2.11. Secondary structure stability correlates with the requirement for *DBP2* in termination of protein-coding genes.

(A) Schematic representation the *YOP1* termination reporter. The *YOP1* 3' UTR region was cloned into the intron of the previously described *CUP1* termination reporter plasmid (Steinmetz and Brow, 2003). (B) Insertion of the *YOP1* 3'UTR into the termination reporter confers *DBP2*-dependent copper sensitivity to *cup1Δ* cells. Serial dilution of *cup1Δ* (top) or *cup1Δ dbp2Δ* cells (bottom) transformed with the reporter without terminator insertion before *CUP1* (no terminator), with wild type *CYC1* terminator (*CYC1 TER*), or with wild type *YOP1* 3' UTR on non-selective media (0 mM) or media containing high concentrations (1.2mM) of copper sulfate (CuSO₄). (C) Predicted secondary structure of the *YOP1* 3'UTR with structure-destabilizing mutations. The secondary structure of the 3'UTR of *YOP1* in *dbp2Δ* was predicted using DMS reactivity patterns and the ViennaRNA package (Lorenz et al., 2011). Adenosines and cytosines with reduced DMS reactivity in *dbp2Δ* are colored in red and those with increased reactivity are in blue. The mutated nucleotides for the reporter assay are indicated by arrows and the folding energy (ΔG) for the wild-type and mutant structures were predicted using ViennaRNA package 2.0 (temperature parameter = 30 °C) (Lorenz et al., 2011). (D) Destabilization of the *YOP1* 3'UTR bypasses requirement for *DBP2* in termination. A liquid growth assay of *cup1Δ* or *cup1Δ dbp2Δ* cells with the reporter without internal terminator insertion (no terminator), with the wild-type *YOP1* 3' UTR, and with the destabilized *YOP1* 3' UTR containing 1 (3'UTR-mut1) or 2 (3'UTR-mut1+mut2) structural mutation sites in the presence or absence of 0.4mM CuSO₄. A reduced amount of CuSO₄ was necessary as compared to panel C as the cells showed enhanced sensitivity in liquid culture as compared to plates.

Table 2.1. Yeast strains. Strains are all isogenic and correspond to the BY4741 or S288C background.

Strain	Genotype	Source/Reference
<i>Wild type (BY4741)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	Open Biosystems
<i>dbp2Δ</i>	<i>MATa dbp2::KanMx6 his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	(Cloutier et al., 2012)
<i>DBP2-3XFLAG</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 DBP2-3XFLAG-KanMx</i>	(Cloutier et al., 2012)
<i>cup1Δ</i>	<i>MATa cup1Δ0 his3Δ0 trp1Δ0 leu2Δ0 ura3Δ0 lys2Δ0 ade2Δ0</i>	(Steinmetz and Brow, 1996)
<i>cup1Δdbp2Δ</i>	<i>MATa dbp2::HygR cup1Δ0 his3Δ0 trp1Δ0 leu2Δ0 ura3Δ0 lys2Δ0 ade2Δ0</i>	This study
<i>NRD1-3XFLAG</i>	<i>MATa his3D1 leu2D0 met15D0 ura3D0 NRD1-3XFLAG-KanMx</i>	This study
<i>NRD1-3XFLAG dbp2Δ</i>	<i>MATa his3D1 leu2D0 met15D0 ura3D0 NRD1-3XFLAG-KanMx dbp2::HygR</i>	This study
<i>RPB3-3XFLAG</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-3XFLAG-KanMx</i>	This study
<i>RPB3-3XFLAG dbp2Δ</i>	<i>MATa his3D1 leu2D0 met15D0 ura3D0 RPB3-3XFLAG-KanMx dbp2::HygR</i>	This study

Table 2.2. Plasmids used in this study. Plasmids were used for endogenous, 3X-FLAG tagging or as the termination reporter (Figure 2.11).

Plasmid	Source/Reference
<i>p3XFLAG:KanMX</i>	(Gelbart et al., 2001)
<i>pGAC24</i>	(Lesser and Guthrie, 1993)
<i>pGAC24-CYC1 terminator</i>	(Steinmetz and Brow, 2003)
<i>pGAC24-YOP1 3'UTR</i>	This study
<i>pGAC24-YOP1 3'UTR-mut1</i>	This study
<i>pGAC24-YOP1 3'UTR-mut1+mut2</i>	This study

Table 2.3. Primers for strain construction and cloning. Primers were used for endogenous, 3X-FLAG tagging or for reporter construction.

<i>NRD1-3XFLAG F</i>	5'- GCTCAATTGAATTCTTTGATGAATATGCTTAACCAACAGCAG CAGCAACAACAACAAAGCAGGGAACAAAAGCTGGAG -3'
<i>NRD1-3XFLAG R</i>	5'- TATATATAGAGGTAGATTAGTTTTATGTACTATGAGCAAATA AAGGGTGGAGTAAAGATCCTATAGGGCGAATTGGGT -3'
<i>YOP1-XhoI-F</i>	5'- ACCGCTCGAGGAGGCAACTCACCTATATCCTC -3'
<i>YOP1-XhoI-R</i>	5'- ATACCTCGAGTAAACGACTCCAAAATATTTTATATGTTAAG - 3'
<i>YOP1-mut F</i>	5'- GTTGTTATATATCGGTACTCAAGATATAAAATATTTTGG A -3'
<i>YOP1-mut R</i>	5'- CGACTCCAAAATATTTTATATCTTGAGTACCGATATATAAC - 3'
<i>YOP1-mut2 F</i>	5'- TTCTCTTTGTTATCTATTCTATTCAGAGAATATAAGTACATAT G -3'
<i>YOP1-mut2 R</i>	5'- CAACTAATAACATATGTACTTATATTCTCTGAATAGAATAGA TAAC -3'

Table 2.4. Sequences used for iCLIP-seq. Bolded letters are barcodes for different libraries.

Biotinylated 3' adaptor	5'-/5rApp/AGATCGGAAGAGCGGTTCAG/3Biotin/-3'
RT primer 1	5'-/5phos/DDDNN AAC NNNNAGATCGGAAGAGCG TCGTGAT/iSp18/GG ATCC/iSp18/TACTGAACCGC-3'
RT primer 2	5'-/5phos/DDDNN ACA NNNNAGATCGGAAGAGCG TCGTGAT/iSp18/GG ATCC/iSp18/TACTGAACCGC-3'
RT primer 3	5'-/5phos/DDDNN ATTG NNNNAGATCGGAAGAGCG TCGTGAT/iSp18/GG ATCC/iSp18/TACTGAACCGC-3'
P3_short	5'-CTGAACCGCTCTTCCGATCT-3'
P5_short	5'-ACACGACGCTCTTCCGATCT-3'
P3_Solexa	5'-CAAGCAGAAGACGGCATACGAGATCGGTCTC GGCATTCTGCTGAACCGCTCTTCCGATCT-3'
P5_Solexa	5'-AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGCTCTTCCGATCT-3'

Table 2.5. Oligonucleotides used for ChIP. Oligos correspond to Figure 2.3 & 2.6.

NRD1-5'-ChIP-F	5'- AGGGCAAGTGTTCGTCC -3'
NRD1-5'-ChIP-R	5'- AAACCTCGTAAAGGGAAGGAGC -3'
NRD1-5'-ChIP-probe	/56-FAM/CCTCCATGT/ZEN/TCCATTCTCGTTAGCA/3IABkFQ/
PCF11-5'-ChIP-F	5'- ATTGGATGAGAACTTGGCCT -3'
PCF11-5'-ChIP-R	5'- CTCCGAAAATTGTCCTGGTAATTC -3'
PCF11-5'-ChIP-probe	/56-FAM/TTTGAAATT/ZEN/TCGCTTTCTTCCCATGCCT/3IABkFQ/
YOP1-ORF-ChIP-F (1)	5'- GGTTTTATTGGTTCCATCGTCATC -3'
YOP1-ORF-ChIP-R (1)	5'- CTTTCTTGTTAGTATAACTGCTTCGG -3'
YOP1-ORF-ChIP-probe (1)	/56-FAM/TGTGTCGGG/Zen/TCATTTGGCTGTTTG/3IABkFQ/
YOP1-3'UTR-ChIP-F (2)	5'- AGACAGAAAAGGATGAAATTAGAGC -3'
YOP1-3'UTR-ChIP-R (2)	5'- TTTGAGGATATAGGTGAGTTGCC -3'
YOP1-3'UTR-ChIP-probe (2)	/56-FAM/CTGTAGCCT/ZEN/TAGAAGCCTCATTGACGG/3IABkFQ/
YOP1-readthrough-ChIP-F (3)	5'- TTGGAGTCGTTTATGGTGTCC -3'
YOP1-readthrough-ChIP-R (3)	5'- TCTGTCGCGCATATCAAGAG -3'
YOP1-readthrough-ChIP-probe (3)	/56-FAM/TCATCGTGT/Zen/TGTGTCGTGACGTGT/3IABkFQ/
RBG1-ORF-ChIP-F (1)	5'- CCAAAAGAACAAGGCCACATC -3'
RBG1-ORF-ChIP-R (1)	5'- GCCACATCAAAACCAATACCAG -3'
RBG1-ORF-ChIP-probe (1)	/56-FAM/TGGGTCAAC/Zen/TGAAGGCCAAGCTG/3IABkFQ/

Table 2.5 continued

RBG1-readthrough-ChIP-F (3)	5'- CTACCGAGGGCTCTTCAAAAG -3'
RBG1-readthrough-ChIP-R (3)	5'- AGATGTGTCAATTTACCAGAAAAGTC -3'
RBG1-readthrough-ChIP-probe (3)	/56-FAM/CGACCTCGC/Zen/TTTGTGGTACCCAT/3IABkFQ/
YNL190W- ORF-ChIP-F (1)	5'- CTATTACTCTAGCCACCGTTGC -3'
YNL190W- ORF-ChIP-R (1)	5'- AACTACCGTCCGATGACAAAG -3'
YNL190W- ORF-ChIP-probe (1)	/56-FAM/TGCCACTGC/Zen/TAAGAAGGGTGAACAT/3IABkFQ/
YNL190W-3'UTR-ChIP-F (2)	5'- GGAAGACCTAATTTTCTCCGGT -3'
YNL190W-3'UTR-ChIP-R (2)	5'- CACAAGCACACGTAAACACATAG -3'
YNL190W-3'UTR-ChIP-probe (2)	/56-FAM/TCAGGATGA/ZEN/AGGGGGTAGGGGG/3IABkFQ/
YNL190W-readthrough-ChIP-F (3)	5'- GAGAAAGATCATCCAACCTTTTAATCATG -3'
YNL190W-readthrough-ChIP-R	5'- AGCGTAGAAATAAGGAAAAAGAGAAAG -3'
YNL190W-readthrough-ChIP-probe	/56-FAM/CGAATGTGG/ZEN/AAGGAAAATAGAGCGGAGC/3IABkFQ/

Table 2.6. Oligonucleotides and PCR primers for Structure-Seq. Bolded letters are barcodes for multiple libraries.

Random-hex RT-primer	5'-CAGACGTGTGCTCTTCCGATCTNNNNNN-3'
ssDNA linker	5'- /5Phos/NNNAGATCGGAAGAGCGTCGTGTAG/3SpC3/- 3'
Illumina TruSeq forward primer	5'-AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCCGATCT-3'
Illumina TruSeq reverse primer_index 1	5'-CAAGCAGAAGACGGCATAACGAGATTGGTCAG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
Illumina TruSeq reverse primer_index 2	5'-CAAGCAGAAGACGGCATAACGAGATGATCTGG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
Illumina TruSeq reverse primer_index 3	5'-CAAGCAGAAGACGGCATAACGATCGTGATG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

Table 2.7. The list of read-through transcripts in *dbp2Δ*. The ratio of 3' extended / total transcripts is also listed for both wild type and *dbp2Δ*.

Gene ID	Gene Name	WT	<i>dbp2Δ</i>
Q0160	SCEI	11.30555556	18.84154316
YAL016W	TPD3	0.149871278	0.471447906
YAL030W	SNC1	0.012197879	0.037071203
YAL037W		0.435803357	1.017882188
YAR003W	SWD1	0.063830182	0.146853002
YBL024W	NCL1	0.015350627	0.062969342
YBL041W	PRE7	0.086855396	0.15901146
YBL071W-A	KTI11	0.008773651	0.03249847
YBL081W		0.075481866	0.395854136
YBL092W	RPL32	0.215687597	0.696376788
YBR048W	RPS11B	0.000546324	0.00394011
YBR056W		0.04851038	0.227005827
YBR111W-A	SUS1	0.010244361	0.029864195
YBR139W		0.425455575	1.473008616
YBR149W	ARA1	0.107032957	0.33793132
YBR165W	UBS1	0.053864427	0.131444891
YBR171W	SEC66	0.028339358	0.085485954
YBR175W	SWD3	0.046903935	0.222674861
YBR189W	RPS9B	0.059583088	0.130786084
YBR194W	AIM4	0.005558771	0.024518508
YBR242W		0.284193664	0.864187024
YBR244W	GPX2	0.001167496	0.01997883
YBR246W	RRT2	0.024877708	0.217394704
YBR271W	EFM2	0.017593003	0.04343686
YBR278W	DPB3	0.01616635	0.037425972
YBR290W	BSD2	0.1561623	0.484105479
YBR295W	PCA1	0.055333426	0.242968847
YCR016W		0.025106599	0.118851365
YCR034W	ELO2	0.033969379	0.071402613
YCR036W	RBK1	0.071111435	0.125131526
YCR045W-A		0.026279894	0.156459898
YCR060W	TAH1	0.016793887	0.216661789
YCR063W	BUD31	0.034070933	0.332931579
YCR066W	RAD18	0.206806722	0.692170973
YCR073W-A	SOL2	0.004958217	0.026383146
YDL004W	ATP16	0.044655431	0.131271075

Table 2.7 continued

YDL006W	PTC1	0.077460262	0.462050992
YDL013W	SLX5	0.007598903	0.042560448
YDL051W	LHP1	0.003815732	0.01227894
YDL066W	IDP1	0.004105087	0.039885932
YDL073W	AHK1	0.034378505	0.129248613
YDL099W	BUG1	0.004697657	0.02438357
YDL102W	POL3	0.05746974	0.220589608
YDL137W	ARF2	0.028988878	0.075951341
YDL155W	CLB3	0.006988037	0.025195133
YDL168W	SFA1	0.005933475	0.041702517
YDL178W	DLD2	0.012402644	0.069484319
YDL212W	SHR3	0.026628892	0.078443649
YDL240W	LRG1	0.003233438	0.032643237
YDR007W	TRP1	0.021417881	0.059151912
YDR021W	FAL1	0.017142292	0.052464181
YDR031W	MIX14	0.039140484	0.245376415
YDR065W	RRG1	0.018698986	0.062299482
YDR147W	EKI1	0.16379955	0.396593354
YDR168W	CDC37	0.103996972	0.31982659
YDR174W	HMO1	1.236411746	4.456076989
YDR214W	AHA1	0.011538712	0.044987353
YDR232W	HEM1	0.160682559	0.340188354
YDR235W	PRP42	0.003818824	0.008198825
YDR242W	AMD2	0.020721503	0.049124581
YDR262W		0.034472125	0.166405736
YDR299W	BFR2	0.007242072	0.028019169
YDR307W	PMT7	0.008353631	0.04070362
YDR318W	MCM21	0.029531633	0.050108475
YDR325W	YCG1	0.022320159	0.080363555
YDR358W	GGA1	0.057366051	0.167734601
YDR363W	ESC2	0.061708667	0.19545142
YDR363W-A	SEM1	0.005585032	0.016628605
YDR373W	FRQ1	0.005244328	0.041132638
YDR377W	ATP17	0.052613532	0.174960546
YDR381W	YRA1	0.017668984	0.110303719
YDR386W	MUS81	0.042007119	0.217412963
YDR389W	SAC7	0.012994683	0.114724599

Table 2.7 continued

YDR427W	RPN9	0.009985393	0.029016028
YDR452W	PPN1	0.019390641	0.085022903
YDR463W	STP1	0.021717209	0.094957567
YDR472W	TRS31	0.027869061	0.057238856
YDR513W	GRX2	0.017268658	0.075554883
YEL018W	EAF5	0.036391228	0.138063651
YEL024W	RIP1	0.110947157	0.208414287
YEL032W	MCM3	0.013573292	0.030552132
YEL043W		0.154127063	0.362797008
YEL044W	IES6	0.026405917	0.058569086
YEL072W	RMD6	0.021373424	0.102703889
YER009W	NTF2	0.270322546	0.631705921
YER016W	BIM1	0.016381566	0.036010706
YER019W	ISC1	0.008334204	0.037517711
YER030W	CHZ1	0.070723981	0.183470696
YER032W	FIR1	0.249416924	0.814527069
YER048W-A	ISD11	0.078161373	0.246943013
YER063W	THO1	0.1192381	0.355491835
YER074W	RPS24A	0.265559972	0.650148689
YER088W-B		0.001919219	0.020285873
YER092W	IES5	0.001949373	0.017601701
YER127W	LCP5	0.04705852	0.157531172
YER128W	VFA1	0.021672948	0.10212438
YER129W	SAK1	0.131389099	0.441143205
YER131W	RPS26B	0.050650406	0.322278702
YER166W	DNF1	0.138068959	0.499870348
YER185W	PUG1	0.001490182	0.028689809
YER188W		0.326581028	1.463894682
YFL049W	SWP82	0.005179779	0.027204803
YFR001W	LOC1	0.011635045	0.034944079
YFR008W	FAR7	0.469281219	1.11507832
YFR012W-A		0.093353006	0.353616602
YFR013W	IOC3	0.000600166	0.008811136
YFR034W-A		0.067221953	0.120714065
YFR052W	RPN12	0.116764313	0.268686616
YFR055W	IRC7	0.546781348	1.251320713
YGL010W	MPO1	0.034997641	0.171858448

Table 2.7 continued

YGL012W	ERG4	0.044159544	0.239385727
YGL041W-A		0.000938623	0.005593305
YGL111W	NSA1	0.180364168	0.594815482
YGL130W	CEG1	0.613756614	3.570393375
YGL155W	CDC43	0.138262868	0.325213647
YGL159W		0.010940649	0.028577403
YGL162W	SUT1	2.12535014	4.894590529
YGL174W	BUD13	0.027965249	0.134360845
YGL191W	COX13	0.057574604	0.248974645
YGL202W	ARO8	0.205507612	1.618715399
YGL210W	YPT32	0.00024403	0.004038086
YGL225W	VRG4	0.025674269	0.078766068
YGR038W	ORM1	0.006965917	0.054042869
YGR046W	TAM41	0.03260621	0.08192676
YGR049W	SCM4	0.011237846	0.040379575
YGR054W		0.195733678	0.460149344
YGR097W	ASK10	0.319852744	0.718319213
YGR108W	CLB1	0.046002991	0.088493068
YGR118W	RPS23A	0.036552092	0.081316787
YGR129W	SYF2	0.015749216	0.069245122
YGR149W		0.027580163	0.15244615
YGR170W	PSD2	0.026342258	0.06649281
YGR275W	RTT102	0.244938918	0.530225843
YHL015W	RPS20	0.018600389	0.104218528
YHL027W	RIM101	0.00814305	0.04314844
YHR002W	LEU5	0.021108425	0.066333516
YHR020W		0.003522261	0.026730801
YHR060W	VMA22	0.210389355	0.459642945
YHR074W	QNS1	0.01820213	0.03869244
YHR076W	PTC7	0.000179319	0.004683263
YHR103W	SBE22	0.106457271	0.437878529
YHR122W	CIA2	0.034545913	0.108500549
YHR148W	IMP3	0.000614251	0.009568815
YHR178W	STB5	0.004539722	0.010746736
YHR197W	RIX1	0.01010101	0.0411979
YHR200W	RPN10	0.007694333	0.01792394
YHR214W-A		0.093396926	0.167683167

Table 2.7 continued

YIL009W	FAA3	0.003306371	0.018125269
YIL010W	DOT5	0.055919657	0.103695409
YIL016W	SNL1	0.014871447	0.027724596
YIL019W	FAF1	0.015555556	0.075428968
YIL092W		0.000875204	0.00561594
YIL097W	FYV10	0.015868257	0.038798785
YIR018W	YAP5	0.016127318	0.056326069
YIR023W	DAL81	0.072568625	0.225144021
YJL001W	PRE3	0.018480871	0.078362515
YJL014W	CCT3	0.081990241	0.169430358
YJL030W	MAD2	0.003200219	0.048896204
YJL046W	AIM22	0.032095986	0.088614603
YJL082W	IML2	0.023912868	0.113986685
YJL098W	SAP185	0.078024681	0.376778857
YJL111W	CCT7	0.008260028	0.029550824
YJL115W	ASF1	0.006972666	0.029313435
YJL126W	NIT2	0.79115353	2.784547152
YJL148W	RPA34	0.016281022	0.07662614
YJL159W	HSP150	0.003376798	0.023009545
YJL174W	KRE9	0.00155956	0.008246258
YJL179W	PFD1	0.011276161	0.053682503
YJL191W	RPS14B	0.053124013	0.144558415
YJR002W	MPP10	0.002754821	0.060263523
YJR032W	CPR7	0.013270637	0.034679728
YJR035W	RAD26	0.014977799	0.048442683
YJR055W	HIT1	0.080994366	0.299882461
YJR064W	CCT5	0.070909717	0.108119118
YJR089W	BIR1	0.008369834	0.045826228
YJR094W-A	RPL43B	0.041309559	0.160022282
YJR099W	YUH1	0.07260616	0.169537046
YJR110W	YMR1	0.015128862	0.054244162
YJR144W	MGM101	0.008335764	0.076894083
YKL002W	DID4	0.022058705	0.045778948
YKL007W	CAP1	0.002606689	0.00807984
YKL015W	PUT3	0.004613077	0.017990947
YKL018W	SWD2	0.040065829	0.232870422
YKL023W		0.603216905	1.572621493

Table 2.7 continued

YKL033W	TTI1	0.101927614	0.349774124
YKL041W	VPS24	0.005741948	0.026635636
YKL058W	TOA2	0.626061893	1.64502025
YKL077W		0.001172735	0.006778695
YKL088W	CAB3	0.00244118	0.041667117
YKL096W-A	CWP2	0.009231437	0.052241551
YKL098W	MTC2	0.005509688	0.024801613
YKL117W	SBA1	0.025444872	0.089899372
YKL156W	RPS27A	0.37827822	0.533525219
YKL180W	RPL17A	0.013699044	0.076280012
YKL181W	PRS1	0.026973485	0.046539114
YKL191W	DPH2	0.007681344	0.026942353
YKL195W	MIA40	0.00032587	0.00262662
YKL204W	EAP1	0.010229209	0.033640561
YKL212W	SAC1	0.048213822	0.121789701
YKR003W	OSH6	0.11245758	0.365690298
YKR030W	GMH1	0.052670453	0.08767738
YKR062W	TFA2	0.009077259	0.042739694
YKR097W	PCK1	0.01774277	0.070202924
YKR106W	GEX2	0.021291549	0.10031925
YLL001W	DNM1	0.006786695	0.028737417
YLL008W	DRS1	0.013194484	0.062808361
YLL061W	MMP1	0.067500633	0.239691585
YLR007W	NSE1	0.015845253	0.053659256
YLR009W	RLP24	0.006897851	0.032660968
YLR011W	LOT6	0.004496228	0.038127683
YLR025W	SNF7	0.010281803	0.063410235
YLR048W	RPS0B	0.002324144	0.018495635
YLR061W	RPL22A	0.001143347	0.003962655
YLR066W	SPC3	0.016866837	0.04831719
YLR086W	SMC4	0.029884858	0.126072092
YLR088W	GAA1	0.009175666	0.03047939
YLR135W	SLX4	0.000356344	0.001588585
YLR143W	DPH6	0.00041624	0.001467671
YLR167W	RPS31	0.003966716	0.013795077
YLR180W	SAM1	0.012432725	0.052744139
YLR206W	ENT2	0.094774366	0.443948857

Table 2.7 continued

YLR220W	CCC1	0.029811596	0.073526337
YLR226W	BUR2	0.001784136	0.012738288
YLR243W	GPN3	0.00244623	0.024689831
YLR264W	RPS28B	0.035093164	0.103789248
YLR299W	ECM38	0.015045069	0.05836545
YLR324W	PEX30	0.035329656	0.164946724
YLR350W	ORM2	0.469753086	0.45713141
YLR353W	BUD8	0.00858454	0.049034934
YLR362W	STE11	0.001511179	0.004995715
YLR386W	VAC14	0.129875296	0.21052689
YLR417W	VPS36	0.006279627	0.031413562
YLR419W		0.001404562	0.005700319
YLR432W	IMD3	0.339517345	1.520197044
YLR456W		0.002098709	0.027169877
YML024W	RPS17A	0.002823761	0.019988537
YML065W	ORC1	0.014719394	0.039631464
YML098W	TAF13	0.012441307	0.051684946
YML108W		0.005149796	0.018350116
YML131W		0.031888995	0.077623457
YMR005W	TAF4	0.730315647	1.065830866
YMR019W	STB4	0.027426815	0.143061604
YMR042W	ARG80	0.003834938	0.021910369
YMR054W	STV1	0.00966901	0.035983026
YMR059W	SEN15	0.016165027	0.099998297
YMR061W	RNA14	0.009737895	0.026888042
YMR079W	SEC14	0.259929799	0.384768303
YMR096W	SNZ1	0.103913791	0.372976173
YMR100W	MUB1	0.014290864	0.057338411
YMR122W-A		0.007929076	0.060795631
YMR158W	MRPS8	0.022491627	0.092637052
YMR192W	GYL1	0.00877193	0.032804502
YMR194W	RPL36A	0.222449237	0.861170187
YMR200W	ROT1	0.086462899	0.216683402
YMR246W	FAA4	0.016383862	0.047683387
YMR255W	GFD1	0.004453732	0.023892063
YMR264W	CUE1	0.07510857	0.226490776
YMR289W	ABZ2	0.125674522	0.325232512

Table 2.7 continued

YMR304W	UBP15	0.00731379	0.069672323
YNL015W	PBI2	0.163581513	0.469483516
YNL038W	GPI15	0.177559301	0.704832105
YNL056W	OCA2	0.006131839	0.064866064
YNL085W	MKT1	0.836212165	1.870390671
YNL094W	APP1	0.002042276	0.017935224
YNL107W	YAF9	0.007362918	0.060196621
YNL162W-A		0.060957229	0.279195402
YNL178W	RPS3	0.016804924	0.075375999
YNL190W		0.030561488	0.086096925
YNL193W		0.020677177	0.176551184
YNL233W	BNI4	0.019642887	0.055193513
YNL289W	PCL1	0.00860797	0.02970912
YNL312W	RFA2	0.011658218	0.055844975
YNL317W	PFS2	0.004893164	0.015855473
YNR010W	CSE2	0.004895734	0.03007154
YNR012W	URK1	0.031348554	0.072040469
YNR021W		0.108529801	0.22313754
YNR058W	BIO3	0.031653559	0.108361781
YNR059W	MNT4	0.016181378	0.162073027
YOL004W	SIN3	0.007251752	0.018009166
YOL008W	COQ10	0.013462931	0.111066842
YOL032W	OPI10	0.005372061	0.02200218
YOL039W	RPP2A	0.023681926	0.156621921
YOL069W	NUF2	0.251633012	0.604236239
YOL090W	MSH2	0.014779004	0.037598397
YOL093W	TRM10	0.074161082	0.170133049
YOL103W	ITR2	0.008797204	0.054659633
YOL125W	TRM13	0.069383203	0.198539518
YOL127W	RPL25	0.306157644	0.799992545
YOL129W	VPS68	0.002250863	0.023289832
YOL144W	NOP8	0.013416131	0.032521117
YOR004W	UTP23	0.182911572	0.538052683
YOR019W		0.115748216	0.35848987
YOR061W	CKA2	0.879365079	2.141037736
YOR063W	RPL3	0.170390396	0.253482967
YOR066W	MSA1	0.005119448	0.020571546

Table 2.7 continued

YOR109W	INP53	0.049803041	0.110403457
YOR142W	LSC1	0.015389933	0.096169316
YOR210W	RPB10	0.01091394	0.086721377
YOR260W	GCD1	0.081059533	0.274939891
YOR291W	YPK9	0.02063751	0.043960048
YOR293W	RPS10A	0.015875586	0.049864707
YOR304W	ISW2	0.006512261	0.022045184
YOR317W	FAA1	0.006976872	0.033490152
YOR341W	RPA190	0.130787339	0.475761004
YPL043W	NOP4	0.168586422	0.562045418
YPL081W	RPS9A	0.002109887	0.01471388
YPL087W	YDC1	0.178050053	0.599736765
YPL129W	TAF14	0.04519774	0.130845747
YPL135W	ISU1	0.011612398	0.029532557
YPL143W	RPL33A	0.025194076	0.092156582
YPL150W		0.034932245	0.096326724
YPL160W	CDC60	0.002304861	0.01981569
YPL175W	SPT14	0.00021858	0.00161956
YPL180W	TCO89	0.031934395	0.096946565
YPL183W-A	RTC6	0.130090722	0.418785757
YPL185W		0.00429937	0.022748996
YPL203W	TPK2	0.006062151	0.016593133
YPL213W	LEA1	0.014174615	0.038371688
YPL228W	CET1	0.000507737	0.003988445
YPL235W	RVB2	0.030994953	0.100359864
YPL260W		0.170796547	0.37392173
YPR001W	CIT3	0.004592933	0.012491696
YPR018W	RLF2	0.023737782	0.097293001
YPR028W	YOP1	0.027636083	0.100232594
YPR043W	RPL43A	0.03754914	0.111622913
YPR094W	RDS3	0.000855191	0.005709743
YPR103W	PRE2	0.058940741	0.216572968
YPR127W		0.251683514	0.45645206
YPR159W	KRE6	0.006133868	0.04793568
YPR165W	RHO1	0.001801388	0.01461845
YPR175W	DPB2	0.081413961	0.215015334
YPR196W		0.000762189	0.006965765

Table 2.7 continued

YAL010C	MDM10	0.017561682	0.047554604
YAL029C	MYO4	0.000750312	0.006147244
YAL036C	RBG1	0.041471158	0.101837501
YAL040C	CLN3	0.013304786	0.04616226
YAL049C	AIM2	0.023194017	0.169814911
YAL067C	SEO1	0.15886208	0.498808669
YAR002C-A	ERP1	0.008960985	0.041481246
YBL010C		0.000171708	0.005674623
YBL014C	RRN6	0.003875969	0.03058774
YBL028C		0.010163381	0.037436954
YBL030C	PET9	0.092501524	0.228138183
YBL055C		0.004799684	0.036312551
YBR001C	NTH2	0.0251825	0.070825924
YBR009C	HHF1	0.01110026	0.056004095
YBR011C	IPP1	0.030838479	0.146291291
YBR014C	GRX7	0.016915579	0.055032114
YBR045C	GIP1	0.091362946	0.445425725
YBR049C	REB1	0.742606943	1.37034632
YBR055C	PRP6	0.122288051	0.402377012
YBR079C	RPG1	0.012284894	0.037398903
YBR082C	UBC4	0.022707032	0.095021962
YBR092C	PHO3	0.037343952	0.247437719
YBR109C	CMD1	0.044786018	0.16976581
YBR138C		0.797403577	2.23120915
YBR143C	SUP45	0.015768412	0.072656806
YBR154C	RPB5	0.086442925	0.275794848
YBR166C	TYR1	0.001239364	0.010491089
YBR181C	RPS6B	0.056765469	0.131488913
YBR216C	YBP1	0.021846888	0.122657898
YBR231C	SWC5	0.060846327	0.208946111
YBR234C	ARC40	0.015877509	0.108086472
YBR240C	THI2	0.001000373	0.005299614
YBR255C-A		0.003616476	0.019262828
YBR260C	RGD1	0.010024122	0.028808062
YBR283C	SSH1	0.038703745	0.205959752
YCL031C	RRP7	0.014280676	0.061111898
YCL035C	GRX1	0.287668437	0.458419493

Table 2.7 continued

YCL037C	SRO9	0.012938333	0.050280977
YCL049C		0.207664867	0.606935125
YCL061C	MRC1	0.102130824	0.264010931
YCR004C	YCP4	0.194444444	0.530926501
YCR027C	RHB1	0.017486676	0.047391972
YCR031C	RPS14A	0.099009745	0.214367837
YCR090C		0.024095285	0.119833989
YCR092C	MSH3	0.055355579	0.143054353
YCR095C	OCA4	0.000780636	0.00443645
YDL005C	MED2	0.00691547	0.017349125
YDL012C		0.71645492	2.175550636
YDL043C	PRP11	0.063633057	0.151636763
YDL045C	FAD1	0.017259432	0.038450532
YDL059C	RAD59	0.012986414	0.105797956
YDL061C	RPS29B	0.024457673	0.070848459
YDL063C	SYO1	0.009703181	0.04208167
YDL072C	YET3	0.033821922	0.113459995
YDL081C	RPP1A	0.012633513	0.047113298
YDL098C	SNU23	0.018106482	0.050179544
YDL101C	DUN1	0.032285086	0.049249559
YDL126C	CDC48	0.094625889	0.461524435
YDL140C	RPO21	0.053980395	0.170534699
YDL145C	COP1	0.484449571	1.009882935
YDL167C	NRP1	0.045498972	0.125640164
YDL184C	RPL41A	0.037813958	0.162779085
YDL203C	ACK1	0.10073152	0.262605897
YDL224C	WHI4	0.035158531	0.117913808
YDL226C	GCS1	0.066069203	0.253382933
YDL243C	AAD4	0.001571919	0.011927935
YDR005C	MAF1	0.019157544	0.075110938
YDR006C	SOK1	1.687540069	4.714697185
YDR032C	PST2	0.013104037	0.039107409
YDR034C	LYS14	0.002710027	0.141058291
YDR050C	TPI1	0.181503088	0.345653489
YDR054C	CDC34	0.003870674	0.030020457
YDR079C-A	TFB5	0.019828816	0.08474837
YDR101C	ARX1	0.004562334	0.013644104

Table 2.7 continued

YDR109C		0.039692609	0.126162461
YDR117C	TMA64	0.005804647	0.012004327
YDR123C	INO2	0.143935062	0.34740918
YDR133C		0.018251956	0.053820362
YDR141C	DOP1	0.017279264	0.048464532
YDR151C	CTH1	0.005371368	0.037915154
YDR179C	CSN9	0.078799357	0.150589102
YDR190C	RVB1	0.001025227	0.003845397
YDR196C	CAB5	0.196163393	0.449002477
YDR224C	HTB1	0.003539907	0.016641363
YDR231C	COX20	0.012743652	0.065927647
YDR238C	SEC26	0.014565555	0.071128077
YDR240C	SNU56	0.096193235	0.365285042
YDR276C	PMP3	0.076637137	0.310199579
YDR277C	MTH1	0.17006002	0.455783203
YDR298C	ATP5	0.005507167	0.063076103
YDR306C		0.00541851	0.047686889
YDR308C	SRB7	0.008947761	0.040540046
YDR333C	RQC1	0.013428122	0.054681908
YDR359C	EAF1	0.00666159	0.041871496
YDR361C	BCP1	0.000905797	0.020574596
YDR369C	XRS2	0.087825359	0.316528507
YDR390C	UBA2	0.001480998	0.011980417
YDR404C	RPB7	0.003424826	0.013547232
YDR422C	SIP1	0.053880735	0.192009365
YDR458C	HEH2	0.002820353	0.010382673
YDR473C	PRP3	3.17E-05	0.00161159
YDR500C	RPL37B	0.375	0.324444444
YEL005C	VAB2	0.01069657	0.075361201
YEL029C	BUD16	0.29282383	0.842700564
YEL037C	RAD23	0.421056021	1.607166409
YEL046C	GLY1	0.138620897	0.285610359
YEL055C	POL5	0.020137823	0.060305857
YER008C	SEC3	0.002975593	0.007086177
YER026C	CHO1	0.024465705	0.13550718
YER036C	ARB1	0.005117106	0.035763978
YER056C	FCY2	4.70E-05	0.000744682

Table 2.7 continued

YER071C	TDA2	0.099746974	0.269267093
YER077C	MRX1	0.055351988	0.162909921
YER099C	PRS2	0.008318948	0.031788664
YER105C	NUP157	0.002155924	0.013696761
YER113C	TMN3	0.070568505	0.172318343
YER122C	GLO3	0.023505904	0.125490196
YER124C	DSE1	0.021641076	0.060548201
YER126C	NSA2	0.012345446	0.028231462
YER137C		0.094937673	0.25045403
YER139C	RTR1	0.248672355	0.589715304
YER142C	MAG1	0.057123598	0.163282794
YER151C	UBP3	0.087995557	0.463251877
YER161C	SPT2	0.00133723	0.00589307
YER174C	GRX4	0.019999257	0.067862431
YER186C		0.078027568	0.240747044
YER188C-A		0.005889748	0.025207135
YFL039C	ACT1	0.020492037	0.272853592
YFR005C	SAD1	0.272905489	0.846031746
YFR011C	MIC19	0.019220379	0.048610781
YFR016C		0.093534722	0.26566685
YFR041C	ERJ5	0.025536522	0.084031985
YFR050C	PRE4	0.015395652	0.075485093
YGL013C	PDR1	0.013445424	0.034240354
YGL038C	OCH1	0.017218645	0.058873307
YGL044C	RNA15	0.023320431	0.094039335
YGL054C	ERV14	0.029248412	0.063227044
YGL070C	RPB9	0.019394295	0.09645595
YGL078C	DBP3	0.272436984	0.50547925
YGL091C	NBP35	0.003006799	0.011393621
YGL094C	PAN2	0.01018461	0.043374661
YGL112C	TAF6	0.000612687	0.003533663
YGL131C	SNT2	0.010421432	0.101424116
YGL136C	MRM2	0.011861952	0.023429315
YGL140C		0.066112238	0.166149256
YGL142C	GPI10	0.015950703	0.06087066
YGL150C	INO80	0.160065211	0.453929903
YGL167C	PMR1	0.000767987	0.010322948

Table 2.7 continued

YGL206C	CHC1	0.071749331	0.114447234
YGL224C	SDT1	0.150002998	0.424673857
YGL226C-A	OST5	0.003189645	0.029628582
YGL246C	RAI1	0.007936508	0.032834765
YGR015C		0.100717353	0.206546142
YGR020C	VMA7	0.014351422	0.033134303
YGR027C	RPS25A	0.003101184	0.012502146
YGR037C	ACB1	0.22465134	0.607282502
YGR071C	ENV11	0.009571645	0.028094324
YGR081C	SLX9	0.011583606	0.038906503
YGR083C	GCD2	0.084833544	0.351484669
YGR119C	NUP57	0.003515206	0.032305924
YGR128C	UTP8	0.044541425	0.123662831
YGR150C	CCM1	0.286620773	0.683223173
YGR158C	MTR3	0.014335445	0.048254457
YGR161C	RTS3	0.066382576	0.238060017
YGR189C	CRH1	0.439287287	2.259553669
YGR192C	TDH3	0.085927226	0.369255764
YGR193C	PDX1	0.024756786	0.074444385
YGR200C	ELP2	0.006540029	0.026137599
YGR213C	RTA1	0.005286458	0.019987816
YGR235C	MIC26	0.022010471	0.075474694
YGR244C	LSC2	0.103770626	0.22172619
YGR253C	PUP2	0.430616309	1.322872838
YGR259C		0.011111111	0.225042301
YGR279C	SCW4	6.54E-05	0.001427388
YGR280C	PXR1	0.036589426	0.12963187
YGR282C	BGL2	0.003401231	0.023284721
YHL009C	YAP3	0.003666088	0.010811729
YHL033C	RPL8A	0.00595235	0.025918369
YHR044C	DOG1	0.003388282	0.012491814
YHR046C	INM1	0.006065903	0.019213892
YHR077C	NMD2	0.369997435	0.498792804
YHR141C	RPL42B	0.005964757	0.023013504
YHR188C	GPI16	0.031184647	0.081527833
YHR193C	EGD2	0.164488694	0.605981619
YIL007C	NAS2	0.006569672	0.021130266

Table 2.7 continued

YIL017C	VID28	0.00923542	0.039210808
YIL027C	EMC5	0.004163498	0.014042017
YIL069C	RPS24B	0.622061497	1.174620938
YIL075C	RPN2	0.159877858	0.216527083
YIL077C		0.015714044	0.034936132
YIL091C	UTP25	0.035787633	0.064007449
YIL096C	BMT5	0.098202388	0.160097488
YIL102C-A		0.024685728	0.058178714
YIL127C	RRT14	0.236049099	0.471622457
YIL133C	RPL16A	0.006509213	0.027863783
YIL157C	COA1	0.238423284	0.528661738
YIR001C	SGN1	0.167008462	0.593671032
YIR030C	DCG1	0.150476675	0.571235875
YIR039C	YPS6	0.251083322	0.627542494
YJL011C	RPC17	0.021357189	0.096447362
YJL024C	APS3	0.175953376	0.481690518
YJL038C	LOH1	0.065219259	0.391680482
YJL072C	PSF2	0.020243258	0.03325977
YJL101C	GSH1	0.014555924	0.05710477
YJL124C	LSM1	0.044497366	0.157153385
YJL136C	RPS21B	0.000833081	0.007440337
YJL139C	YUR1	0.006432017	0.024119158
YJL147C	MRX5	0.073041719	0.157799217
YJL151C	SNA3	0.073300324	0.169137896
YJR049C	UTR1	0.061354649	0.151694009
YJR065C	ARP3	0.082006714	0.205900326
YJR067C	YAE1	0.024943434	0.060551637
YJR072C	NPA3	0.004409705	0.011501914
YJR076C	CDC11	0.000248116	0.002508969
YJR088C	EMC2	0.003028248	0.016662599
YJR100C	AIM25	0.004526233	0.016085784
YJR161C	COS5	0.096837945	0.16038711
YKL005C	BYE1	0.110895422	0.273519537
YKL006C-A	SFT1	0.000673401	0.012186698
YKL011C	CCE1	0.022267597	0.043023845
YKL013C	ARC19	14.6031746	7.9625
YKL017C	HCS1	0.156610965	0.823755409

Table 2.7 continued

YKL040C	NFU1	0.074017957	0.26760061
YKL046C	DCW1	0.053303714	0.152589474
YKL100C	YPF1	0.03816092	0.136150687
YKL128C	PMU1	0.306048339	0.744328753
YKL151C		0.073102159	1.375454119
YKL159C	RCN1	0.03176305	0.065581493
YKL164C	PIR1	0.079296984	0.28314521
YKL174C	TPO5	0.029332936	0.079317093
YKL179C	COY1	0.023327823	0.086126668
YKL186C	MTR2	0.003897141	0.031045215
YKL197C	PEX1	0.033781881	0.123556897
YKL206C	ADD66	0.000536226	0.002560352
YKL209C	STE6	0.042725042	0.162826346
YKR005C		0.03828243	0.135430612
YKR022C	NTR2	0.048970283	0.107540404
YKR043C	SHB17	0.001918347	0.00866415
YKR071C	DRE2	0.010534973	0.044664405
YKR092C	SRP40	0.018951661	0.048209342
YKR094C	RPL40B	0.00543606	0.017456085
YLL010C	PSR1	0.000609385	0.007849573
YLL013C	PUF3	0.548000611	5.963675214
YLL018C-A	COX19	0.012090496	0.025023123
YLR016C	PML1	0.126813886	0.344539275
YLR023C	IZH3	0.002900502	0.019925892
YLR055C	SPT8	0.013713004	0.064416846
YLR078C	BOS1	0.077461787	0.17172525
YLR082C	SRL2	0.008487239	0.016555418
YLR093C	NYV1	0.066521423	0.237985148
YLR110C	CCW12	0.039094887	0.455173383
YLR117C	CLF1	0.01970392	0.087551761
YLR120C	YPS1	0.127272727	0.443191312
YLR130C	ZRT2	0.054246453	0.148163268
YLR139C	SLS1	0.009025025	0.063214789
YLR147C	SMD3	0.076550792	0.315316172
YLR163C	MAS1	0.003065189	0.028501646
YLR192C	HCR1	0.064008189	0.255264942
YLR193C	UPS1	0.239420412	0.687625719

Table 2.7 continued

YLR215C	CDC123	0.003646863	0.013871269
YLR221C	RSA3	0.001713085	0.004013662
YLR239C	LIP2	0.003300227	0.027263526
YLR262C-A	TMA7	0.006884211	0.026370022
YLR287C		0.137405745	0.574321458
YLR298C	YHC1	0.013845348	0.03291438
YLR316C	TAD3	0.003004321	0.018603955
YLR321C	SFH1	0.006435572	0.025210755
YLR325C	RPL38	0.016776133	0.046585825
YLR333C	RPS25B	0.409572318	0.618925776
YLR354C	TAL1	0.320755727	0.599949864
YLR361C	DCR2	0.694201227	1.216393168
YLR363C	NMD4	0.080240544	0.335125448
YLR370C	ARC18	0.003002595	0.020217179
YLR387C	REH1	0.079029824	0.161515287
YLR395C	COX8	0.038103703	0.111546368
YLR401C	DUS3	0.123309363	0.451532091
YLR406C	RPL31B	0.310659486	0.87459091
YLR421C	RPN13	0.024602876	0.11473334
YLR431C	ATG23	0.002056802	0.016272435
YLR452C	SST2	0.208585848	0.609505541
YML004C	GLO1	0.000262467	0.004005194
YML011C	RAD33	0.25570538	0.433245271
YML026C	RPS18B	0.039710916	0.141519752
YML043C	RRN11	0.040092148	0.105022757
YML073C	RPL6A	0.145474464	0.307690588
YML097C	VPS9	0.025231498	0.09182237
YML107C	PML39	0.001296605	0.007587187
YML110C	COQ5	0.053137943	0.22860955
YML115C	VAN1	0.01317433	0.047199119
YML120C	NDI1	0.01660387	0.080161035
YMR067C	UBX4	0.022066375	0.060319886
YMR097C	MTG1	0.174911266	0.547695358
YMR099C		0.011108932	0.052174419
YMR126C	DLT1	0.021577295	0.09266555
YMR166C	MME1	0.013886097	0.040464987
YMR168C	CEP3	0.047370454	0.108513271

Table 2.7 continued

YMR187C		0.017693565	0.046642073
YMR197C	VTI1	0.112773805	0.235020417
YMR201C	RAD14	0.037829403	0.110636998
YMR207C	HFA1	0.086722463	0.32963859
YMR216C	SKY1	0.182520689	0.530672183
YMR229C	RRP5	0.244737957	0.679106162
YMR235C	RNA1	0.000480592	0.001375121
YMR244C-A	COA6	0.21627078	0.536173473
YMR299C	DYN3	0.130635693	0.343776951
YNL003C	PET8	0.044748422	0.109391012
YNL005C	MRP7	0.006373993	0.040941875
YNL023C	FAP1	0.035897024	0.091574079
YNL024C-A	KSH1	0.093876088	0.296057875
YNL037C	IDH1	0.009287674	0.035671432
YNL047C	SLM2	0.003937663	0.026145256
YNL055C	POR1	0.020529277	0.070419452
YNL062C	GCD10	0.046837532	0.181801545
YNL064C	YDJ1	0.079724883	0.275292747
YNL069C	RPL16B	0.054163017	0.114498228
YNL074C	MLF3	0.000850908	0.009088677
YNL084C	END3	0.025164009	0.089248385
YNL104C	LEU4	0.000859717	0.012433301
YNL115C		0.117343653	0.254913004
YNL118C	DCP2	0.031126607	0.083316127
YNL133C	FYV6	0.009586868	0.16317085
YNL156C	NSG2	0.000504587	0.004429236
YNL159C	ASI2	0.058520888	0.170676104
YNL164C	IBD2	0.043566006	0.155953271
YNL182C	IPI3	1.001916209	4.700396825
YNL196C	SLZ1	0.025624989	0.094079329
YNL221C	POP1	0.024972421	0.06464853
YNL255C	GIS2	0.016148349	0.057341684
YNL305C	BXI1	0.040311879	0.082443606
YNL313C	EMW1	0.004345827	0.019286496
YNL322C	KRE1	0.005864362	0.018209394
YNR011C	PRP2	0.089325309	0.226217786
YNR013C	PHO91	0.133745976	0.454218809

Table 2.7 continued

YNR029C		0.042891359	0.176056991
YNR032C-A	HUB1	0.136829983	0.321621883
YNR053C	NOG2	0.013608809	0.062581318
YNR055C	HOL1	0.000299491	0.011476984
YNR061C		0.043350127	0.172707561
YOL013C	HRD1	0.026523255	0.101168911
YOL018C	TLG2	0.02114252	0.053336314
YOL021C	DIS3	0.022786929	0.055667703
YOL040C	RPS15	0.006084527	0.058314242
YOL060C	MAM3	0.037311761	0.091777483
YOL070C	NBA1	0.21715392	0.435048484
YOL077C	BRX1	0.005524333	0.012956205
YOL101C	IZH4	0.004301075	0.032491403
YOL114C	PTH4	0.127943707	0.347411453
YOL124C	TRM11	0.00153437	0.012560231
YOL128C	YGK3	0.006141195	0.031987128
YOL159C-A		0.047703944	0.10648405
YOR028C	CIN5	0.013348959	0.124564197
YOR038C	HIR2	0.315178883	0.830507403
YOR046C	DBP5	0.408349532	1.605069251
YOR048C	RAT1	0.07105845	0.249244891
YOR052C	TMC1	0.01791198	0.051950535
YOR074C	CDC21	0.028960094	0.067311035
YOR076C	SKI7	0.011862171	0.031017765
YOR081C	TGL5	0.000498089	0.003293959
YOR095C	RKI1	0.109562565	0.593020851
YOR100C	CRC1	0.085882468	0.351749969
YOR103C	OST2	0.005471302	0.023412975
YOR128C	ADE2	0.12135269	0.421985461
YOR129C	AFI1	0.03456853	0.134853465
YOR143C	THI80	0.018268566	0.04888912
YOR164C	GET4	0.124461194	0.369113674
YOR179C	SYC1	0.117920597	0.463535185
YOR207C	RET1	0.01343727	0.060278568
YOR209C	NPT1	0.009169578	0.048817033
YOR226C	ISU2	0.025670072	0.068770328
YOR245C	DGA1	0.099739475	0.340462659

Table 2.7 continued

YOR256C	TRE2	0.003236973	0.010085327
YOR259C	RPT4	0.050777284	0.152774594
YOR261C	RPN8	0.045416239	0.078967389
YOR270C	VPH1	0.003694566	0.02432946
YOR275C	RIM20	0.003125892	0.006675619
YOR279C	RFM1	0.080448201	0.202396871
YOR281C	PLP2	0.180719197	0.627918013
YOR290C	SNF2	0.010565974	0.028318688
YOR298C-A	MBF1	0.094447694	0.290723195
YOR304C-A	BIL1	0.009681557	0.056940688
YOR327C	SNC2	0.098768003	0.393973305
YOR344C	TYE7	0.015808184	0.073831565
YOR357C	SNX3	0.01884437	0.03972271
YOR369C	RPS12	1.355555556	1.069047619
YOR372C	NDD1	0.004622181	0.019324268
YPL002C	SNF8	0.002597335	0.02049375
YPL007C	TFC8	0.213734395	0.884978136
YPL011C	TAF3	0.01532069	0.038062098
YPL013C	MRPS16	0.07730535	0.191303944
YPL015C	HST2	0.05957738	0.161375949
YPL042C	SSN3	0.040642642	0.103884158
YPL053C	KTR6	0.019374048	0.062622931
YPL064C	CWC27	0.760348339	1.296721249
YPL069C	BTS1	0.004066409	0.035214918
YPL078C	ATP4	0.004260147	0.011725534
YPL083C	SEN54	0.175738038	0.420899574
YPL106C	SSE1	0.026401965	0.141075829
YPL115C	BEM3	0.020742013	0.119817024
YPL127C	HHO1	0.038222151	0.140604363
YPL137C	GIP3	7.27E-05	0.002447289
YPL140C	MKK2	0.000213046	0.003340246
YPL146C	NOP53	0.002898551	0.010526573
YPL151C	PRP46	0.049705657	0.186182508
YPL176C	TRE1	0.002886167	0.012676643
YPL186C	UIP4	0.034781798	0.143827109
YPL199C		0.057903806	0.198758351
YPL202C	AFT2	0.17977678	0.410103576

Table 2.7 continued

YPL206C	PGC1	0.037077375	0.084003365
YPL217C	BMS1	0.076092181	0.270401149
YPL234C	VMA11	0.031109458	0.170681449
YPL264C		0.059787228	0.149759511
YPR010C	RPA135	0.273683679	0.735462621
YPR022C	SDD4	0.094075663	0.245140145
YPR029C	APL4	0.008322984	0.03893036
YPR049C	ATG11	0.006918915	0.037491805
YPR085C	ASA1	0.012302531	0.055348223
YPR088C	SRP54	0.023379212	0.170583581
YPR107C	YTH1	0.125340757	0.282052936
YPR112C	MRD1	0.054487827	0.254437486
YPR144C	NOC4	0.015205092	0.062231901
YPR161C	SGV1	0.016837501	0.036540014
YPR179C	HDA3	0.014570825	0.073667097
YPR186C	PZF1	0.516666667	2.67394095
YPR190C	RPC82	0.02230625	0.063596203
YBL039W-B		0.004400166	0.031694008
YBR056W-A		0.049901143	0.140535714
YBR114W	RAD16	0.028276907	0.08013501
YDR033W	MRH1	0.00440044	0.0349103
YDR055W	PST1	0.016048007	0.062475137
YDR077W	SED1	0.236128221	0.66093789
YDR380W	ARO10	0.043976697	0.11566897
YDR489W	SLD5	0.091015161	0.258744484
YFL055W	AGP3	0.010823847	0.040726085
YFR022W	ROG3	0.049324187	0.095889506
YGL045W	RIM8	0.001650915	0.009311732
YGR198W	YPP1	0.032604485	0.112520141
YGR248W	SOL4	0.015234876	0.033957818
YGR271W	SLH1	0.003715545	0.014883988
YHL048W	COS8	0.029194395	0.068167902
YHR023W	MYO1	0.088958441	0.259077927
YHR137W	ARO9	0.001930001	0.007277264
YIL122W	POG1	0.016003309	0.02854174
YIL136W	OM45	0.003921145	0.021523662
YJL092W	SRS2	0.110027946	0.208940124

Table 2.7 continued

YJR092W	BUD4	0.322842242	0.777788862
YJR096W		0.478734602	1.579810705
YJR152W	DAL5	0.003581454	0.016674286
YKL150W	MCR1	0.003588181	0.028185018
YLR174W	IDP2	0.072454476	0.127188212
YLR303W	MET17	0.858962945	2.220768384
YML042W	CAT2	0.021101195	0.069790403
YMR189W	GCV2	0.040598992	0.142063114
YMR251W-A	HOR7	0.228397893	0.411474541
YNL093W	YPT53	0.066135388	0.208162428
YNL160W	YGP1	0.130793029	0.253843018
YOL140W	ARG8	0.007985598	0.0254751
YOR220W	RCN2	0.047177775	0.096373056
YPL024W	RMI1	0.003829984	0.009560061
YPR111W	DBF20	0.030583077	0.129082443
YAR018C	KIN3	0.005067055	0.014374885
YBL045C	COR1	0.070163128	0.198785597
YBR085C-A		0.011505954	0.050134967
YBR302C	COS2	0.005179973	0.018899877
YDL187C		0.255365789	0.485033578
YDR529C	QCR7	0.000677851	0.012801482
YDR533C	HSP31	0.024098401	0.08383949
YER033C	ZRG8	0.009125708	0.039568394
YER054C	GIP2	0.000263905	0.004222434
YER172C	BRR2	0.019836308	0.061298386
YER184C	TOG1	0.00909469	0.019081425
YGR043C	NQM1	0.503498508	1.235407759
YGR086C	PIL1	0.007524536	0.013654555
YGR289C	MAL11	0.092453411	0.601412684
YHR003C	TCD1	0.01020506	0.037391508
YIR031C	DAL7	0.007335924	0.037945127
YJL165C	HAL5	0.036870911	0.108776751
YLR254C	NDL1	0.004801878	0.022323288
YML099C	ARG81	0.03268265	0.121166621
YMR077C	VPS20	0.107471541	0.43720282
YMR145C	NDE1	0.070099375	0.274248982
YNL058C		0.043571428	0.079420562

Table 2.7 continued

YNL125C	ESBP6	0.133342642	0.402413031
YNL331C	AAD14	0.004787785	0.024053317
YOL052C-A	DDR2	0.025468544	0.109029842
YOL119C	MCH4	0.00646305	0.02729526
YOL136C	PFK27	0.012371824	0.049298171
YOR058C	ASE1	0.004540172	0.020619418
YOR161C	PNS1	0.005051702	0.031489238
YPL004C	LSP1	0.014099901	0.061474225
YPL089C	RLM1	0.024313046	0.064434521
YPL159C	PET20	0.05739528	0.173164304
YPL177C	CUP9	0.0023098	0.013875625

Table 2.8. The contingency table for a Fisher's exact test of the correlation between the list of transcripts bound by Dbp2 at the 3' end and transcripts with read-through defects in *dbp2*Δ (related to Figure 2.5)

		Read-through in <i>dbp2</i> Δ		
		Yes	No	Total
Dbp2-bound at 3' end	Yes	222	497	719
	No	602	2107	2709
	Total	824	2604	3428

Table 2.9. The list of transcripts with significant *DBP2*-dependent structural changes.

YOR153W	YPL263C	YMR116C	YKL143W	YPL081W	YGR175C
YJR009C	YOR271C	YDL081C	YGL148W	YBR025C	YHR179W
YKL054C	YHR027C	YDR276C	YNL121C	YPL037C	YAL012W
YCR012W	YDL161W	YIL053W	YPL249C-A	YKL082C	YDR133C
YOR383C	YHR019C	YKL081W	YGR145W	YOR157C	YNL007C
YHL033C	YPL217C	YHR203C	YPR103W	YJL159W	YOR164C
YLR249W	YLR060W	YLR150W	YGR245C	YDR077W	YDL084W
YPR035W	YLR175W	YDR190C	YDL065C	YOL088C	YOR209C
YMR186W	YDL126C	YGR192C	YHR193C	YGL123W	YMR318C
YLL024C	YGR211W	YPL019C	YGR103W	YIL133C	YDL099W
YPL106C	YKL060C	YGR279C	YDL148C	YNL307C	YDL014W
YPL160W	YHR089C	YGL200C	YJL115W	YOL077C	YOR007C
YMR120C	YPL043W	YGR159C	YPR163C	YPR100W	YLL008W
YHR020W	YKL029C	YGR208W	YMR246W	YDR341C	YBR164C
YJR045C	YOR375C	YEL037C	YMR108W	YER002W	YBR221C
YGL253W	YBL076C	YFR037C	YLR172C	YER133W	YDR381W
YDL029W	YGR189C	YIL052C	YPL094C	YLR262C-A	YER031C
YPL198W	YDL055C	YOR254C	YJR105W	YHR064C	YFR031C-A
YER036C	YML069W	YER055C	YAL036C	YDL226C	YLR048W
YOR361C	YAL003W	YGL245W	YJL124C	YMR290C	YLR350W
YPL240C	YBR249C	YML074C	YNR053C	YPL210C	YOL144W
YPR145W	YPR112C	YBR189W	YLR354C	YNL079C	YER177W
YGR234W	YGR162W	YFL039C	YIL078W	YHL034C	YER025W
YDL153C	YDR346C	YAR015W	YML063W	YGR254W	YOR298C-A
YLR355C	YKR001C	YDL229W	YDR427W	YNR021W	YOR234C
YGL009C	YOR052C	YLR192C	YMR217W	YPR062W	YEL002C
YNL308C	YLR300W	YNL166C	YKR092C	YFR004W	YNL302C
YLR058C	YLR259C	YOR362C	YCL064C	YER048C	YKL154W
YKL164C	YPR165W	YER026C	YOL086C	YOR281C	YDR037W
YAL038W	YPR074C	YAL042W	YJR072C	YHR146W	YFL038C

Table 2.9 continued

YDR361C	YFL010C	YDR394W	YHR025W	YPR080W	YAL025C
YJL189W	YOR259C	YLR333C	YMR307W	YOL041C	YKR081C
YMR002W	YLR449W	YLR448W	YJR076C	YER126C	YDR353W
YDR496C	YNL178W	YKL152C	YGL106W	YKR025W	YGR035C
YER131W	YDR454C	YGL029W	YMR014W	YMR311C	YJL104W
YPL127C	YBR011C	YML073C	YDR050C	YLL045C	YIL118W
YOL111C	YLR028C	YGL221C	YIL051C	YJL158C	YPL225W
YDL020C	YNL247W	YOR232W	YLR018C	YNL208W	YJR123W
YLR429W	YLR441C	YLL018C	YDL147W	YLR147C	YOL040C
YNL281W	YDR210W	YJL140W	YLR109W	YGR180C	YDR224C
YPL204W	YPL146C	YLR340W	YMR043W	YHR026W	YLR065C
YBL002W	YJR002W	YDR120C	YGR282C	YIR026C	YKL016C
YOR206W	YOR272W	YKL013C	YNL301C	YDR226W	YKL180W
YGR229C	YHR010W	YNR051C	YPL090C	YGL078C	YML125C
YPL145C	YGL037C	YER092W	YLR009W	YJL190C	YML026C
YKL216W	YGR020C	YJL123C	YNL175C	YPR188C	YPL004C
YEL001C	YMR194W	YLR029C	YGR167W	YIL018W	YOL127W
YKL192C	YKL099C	YER178W	YEL054C	YIL127C	YPL010W
YMR143W	YJL177W	YFL026W	YPR182W	YMR235C	YPL143W
YPL218W	YPR110C	YGR135W	YCR035C	YGL103W	YGL048C
YDR002W	YPR143W	YOR340C	YLR447C	YOL093W	YLR287C-A
YDR450W	YFR050C	YKL033W-A	YGL012W	YIL069C	YLR395C
YGL105W	YLR276C	YDR098C	YOR293W	YMR260C	YNL069C
YOL120C	YHR152W	YGL135W	YER057C	YDR471W	YNL050C
YCL050C	YOR230W	YGR155W	YOL097C	YKR071C	YHR066W
YBL041W	YBR290W	YDR465C	YGR214W	YBL003C	YCL009C
YOR091W	YKR042W	YGR027C	YEL009C	YNR054C	YKL160W
YNL031C	YBR014C	YMR146C	YLR293C	YDR167W	YEL046C
YDL208W	YNL002C	YNL064C	YOL143C	YJL148W	YER030W
YDR092W	YGR253C	YJR104C	YJL192C	YGR083C	YDR363W-A

Table 2.9 continued

YPL013C	YNL162W	YER019C-A	YBR109C	YFR001W	YEL026W
YJL145W	YPL181W	YLR200W	YMR269W	YDL184C	YDR378C
YKL117W	YMR005W	YLR185W	YER094C	YBR010W	YCR087C-A
YGR148C	YNL112W	YNL156C	YLR390W-A	YNL004W	YKR094C
YCR031C	YDR328C	YLL014W	YKR013W	YDL100C	YBR084C-A
YNL113W	YLR061W	YBR247C	YPR051W	YGL226C-A	YDL002C
YOR189W	YBR143C	YLR167W	YLR022C	YLR421C	YJL136C
YOR224C	YFR003C	YDR064W	YDL051W	YOL016C	YNL255C
YOR369C	YFR052W	YGR086C	YDR233C	YPR187W	YPL129W
YDL143W	YER120W	YOL109W	YDR012W	YPR132W	YHR200W
YBR111C	YLR435W	YPR102C	YER117W	YAR002C-A	YKR057W
YHR062C	YER159C	YBR154C	YBL092W	YDR225W	YER127W
YGR034W	YDR032C	YJR145C	YGL070C	YGR005C	YGR142W
YER100W	YMR072W	YKR035W-A	YML024W	YJR085C	YLR344W
YNL248C	YDR163W	YGL020C	YBR016W	YCR028C-A	YCL035C
YER072W	YPL059W	YNL300W	YPL078C	YLR221C	YBR279W
YMR122W-A	YER007C-A	YBR048W	YJR144W	YBR269C	YER009W
YJL191W	YKL156W	YNL215W	YNL155W	YDL192W	YGR081C
YNL209W	YDL191W	YDR299W	YIL065C	YDL072C	YMR202W
YPR075C	YLR110C	YJR094W-A	YOR115C	YMR242C	YOR287C
YGR106C	YLR243W	YNL110C	YGL030W	YMR038C	YLR068W
YKL018C-A	YMR039C	YPL154C	YOR252W	YOL022C	YDL097C
YDL082W	YHR143W-A	YPR016C	YMR183C	YDL007W	YOR327C
YDR469W	YOR261C	YER102W	YDL166C	YER122C	YLR075W
YNL030W	YJL122W	YLR438C-A	YOR226C	YKR026C	YPR133W-A
YHR005C-A	YIL091C	YDR152W	YBR009C	YHR003C	YOR194C
YBL087C	YMR236W	YOL039W	YDR382W	YIL002W-A	YOR096W
YMR295C	YHR065C	YDR280W	YGR161C	YNL075W	YOL133W
YOL010W	YOR312C	YDL012C	YER074W	YKL006W	YBL001C
YDL130W	YKL041W	YIL008W	YOR063W	YPR028W	YJL179W

Table 2.9 continued

YDL167C	YER050C	YPL117C
YPL079W	YDR071C	YOR265W
YDL083C	YLR325C	YDR412W
YDR139C	YNL015W	YPR148C
YPL220W	YPL243W	YML022W
YLR257W	YKL122C	YBR181C
YOR276W	YBL027W	YBR162W-A
YDR447C	YKL096W-A	YMR123W
YGL054C	YDR121W	YOR004W
YFR028C	YER112W	YCL058W-A
YDR529C	YPL199C	YDR424C
YNL036W	YHR021C	YDR461C-A
YGR118W	YLR388W	
YDL213C	YDL121C	
YBR230C	YNL190W	
YEL027W	YPL064C	
YBR062C	YLL050C	
YCL031C	YNL056W	
YGR280C	YLR051C	
YDL136W	YDR418W	
YHR170W	YOR046C	
YOR182C	YML053C	
YLR406C	YER056C-A	
YGR207C	YGL147C	
YKL009W	YDL061C	
YPR063C	YOR163W	
YNL151C	YGL011C	
YBR106W	YDR461W	
YKR074W	YER146W	
YLR264W	YOL139C	

Table 2.10. The contingency table for a Fisher's exact test of the correlation between the list of Dbp2-bound transcripts and transcripts with significant *DBP2*-dependent structural changes

		Significant <i>DBP2</i> -dependent structural changes		
		Yes	No	Total
Dbp2-bound	Yes	468	1107	1575
	No	144	2962	3106
	Total	612	4069	4681

Table 2.11. The contingency table for a Fisher's exact test of the correlation between the list of transcripts with read-through defects in *dbp2* Δ and transcripts with significant *DBP2*-dependent structural changes (related to Figure 2.10)

		Significant <i>DBP2</i> -dependent structural changes		
		Yes	No	Total
Read-through in <i>dbp2</i> Δ	Yes	180	644	824
	No	321	2283	2604
	Total	501	2927	3428

CHAPTER 3. UNPUBLISHED RESULTS AND FUTURE WORK

3.1 Dbp2 binding is correlated with RNAPII pausing at the 5' end during transcription of protein-coding genes

Previously, it has been shown that RNAPII pauses before termination (1, 2). Consistently, our analysis of RNAPII profiles in wild type cells revealed accumulation of RNAPII before termination sites of snoRNAs (Chapter 2, Figure 2.2B) and around the polyadenylation sites of mRNAs (within ± 200 bp, Chapter 2, Figure 2.5A), and loss of Dbp2 shifted the peak towards downstream, indicating a defect in termination. The promoter-proximal pausing of RNAPII can also be followed by premature termination of transcription (3). Intriguingly, by comparing RNAPII profiles in *dbp2* Δ and wild type at individual protein-coding genes, we observed a reduced level of RNAPII accumulation at 5' ends in the absence of *DBP2* (Chapter 2, Figure 2.6A). We then asked if this lack of RNAPII peak in *dbp2* Δ can be observed in meta-analysis that includes all protein-coding genes. Strikingly, RNAPII pausing at 5' ends is somehow dependent on *DBP2*, as RNAPII signal is reduced in the proximity of the translational start site in *dbp2* Δ as compared to the wild type (Figure 3.1). Additionally, in the absence of *DBP2*, RNAPII signal after 200 bp downstream of the translational start site does not drop as dramatically as observed in the wild type; instead, the change in RNAPII distribution is relatively mild within 500 bp (Figure 3.1). This suggests that RNAPII does not pause normally at the 5' ends of protein-coding genes in cells lacking *DBP2*. Moreover, Dbp2 binding sites near the translational start site also align with RNAPII peak in wild type (Figure 3.1), suggestive of a role of Dbp2 in RNAPII pausing in the beginning of transcription.

3.2 Dbp2 binding at 5' end of mRNAs may be involved in premature termination by NNS

NNS complex has been shown to prematurely terminate multiple protein-coding genes to regulate gene expression (4–6). It has been reported that Dbp2 physically interacts with Sen1 in the NNS complex (7). We also found that Dbp2 promotes efficient recruitment of Nrd1 in the NNS complex to the 5' end of genes potentially targeted by premature termination (Chapter 2, Figure 2.3E). Since premature termination can be a mechanism that regulates gene expression, and the 5' binding of Dbp2 in mRNAs may be functional in premature termination, we then asked if Dbp2

and Nrd1 binding pattern differs in mRNAs that are differentially expressed in *dbp2Δ*. mRNAs were analyzed in three groups: up-regulated, down-regulated, or no change in expression in the absence of *DBP2* (Figure 3.2). Interestingly, transcripts that are up-regulated in *dbp2Δ* display a 5' cluster of Dbp2 and Nrd1 binding, and the 3' peak observed in meta-analysis with all mRNAs is diminished in this specific group (Figure 3.2). This observation leads to a hypothesis that Dbp2 functions in premature termination with NNS complex and repress expression of a subset of genes. This is also consistent with the RNAPII pausing at the 5' ends of protein-coding genes that are impaired in the absence of *DBP2*, potentially due to reduced level of premature termination.

3.3 Future work for study of Dbp2 in premature termination and promoter-proximal RNAPII pausing

3.3.1 To Detect Prematurely Terminated Transcripts by Sequencing of Nascent RNAs in the Presence and Absence of *DBP2*

Genome-wide methods have been developed to capture newly synthesized nascent RNAs, providing a snapshot of cellular transcription (8, 9). This is especially helpful in detection of unstable transcripts (8), including prematurely terminated mRNAs. Briefly, yeast cells are incubated with 4-thiouracil (4tU), an analog of natural uracil. These artificial nucleotides will then be incorporated during a specific timeframe of transcription. After extraction of RNAs, incorporated 4tU can be labeled with biotin and purified using streptavidin-coated magnetic beads. The purified RNA will then be subjected to library preparation and deep sequencing. By comparing 4tU-seq data from *dbp2Δ* and wild-type cells, we will be able to determine whether loss of *DBP2* results in decreased level of prematurely terminated transcripts.

3.3.2 To Determine How Dbp2 is Recruited to the 5' ends of mRNAs

3.3.2.1 Identification of sequence motifs enriched in Dbp2 binding regions at 5' ends of mRNAs

Previously, using HOMER (10), we identified enriched motifs that resemble sequences recognized by Nrd1 and Nab3 in Dbp2 binding regions within 3' UTRs (Chapter 2, Figure 2.3C). This is consistent with the physical interaction between Dbp2 and Sen1. Similarly, HOMER can be used to test if there is any sequence motif preferentially targeted by Dbp2 at 5' ends of mRNAs. This could provide information about the specificity of Dbp2 at the 5' ends and may pinpoint candidate RNA binding proteins interacting with Dbp2.

3.3.2.2 *To test if the recruitment of Dbp2 is dependent on post-translational modifications in the RNAPII C-terminal domain*

The C-terminal domain (CTD) of RNAPII is composed of conserved heptapeptide repeats of the sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser. Post-translational modifications on these amino acids can mark different states of transcription by modulating the interaction between transcription machinery and other proteins (11). Among these modification, Ser2, Thr4, and Tyr1 phosphorylation have been shown to be critical for proper transcription termination of RNAPII transcripts (11, 12). It has also been found that Dbp2 preferentially interacts RNAPII with Thr4 and Tyr1 phosphorylated in CTD compared to RNAPII at other states (13), consistent with a role of Dbp2 in transcription termination.

It was proposed that the recruitment of NNS components may be dependent on the interaction between NNS components and the C-terminal domain (CTD) of RNAPII, and/or the binding of Nrd1 and Nab3 to specific sequence motifs (14, 15). Recently, the phosphorylation of Tyr1 on RNAPII CTD has been reported to be essential for the termination of snoRNAs and a small subset of protein-coding genes, as well as for RNAPII pausing at the 5' end of protein-coding genes (12). Since snoRNAs are mainly terminated through NNS pathway (16), it is possible that Tyr1-phosphorylated CTD has a role in recruitment of NNS components. However, the recruitment of NNS was shown to be independent of their interaction with RNAPII, as mutation of the CTD interacting domain in Nrd1 and Sen1 does not abolish NNS recruitment to their target genes (12). Thus, it remains elusive how premature termination by NNS is regulated.

Taken together, I hypothesize that the recruitment NNS components may be dependent on Dbp2, and Dbp2 can be recruited by interacting with Tyr1 at the 5' ends of protein coding genes, facilitating RNAPII pausing and premature termination. This can be tested by analyzing Dbp2 recruitment at 5' ends of genes such as *NRD1* and *PCF11* in wild type and in mutant with Tyr1 mutation in RNAPII CTD by CHIP of FLAG-tagged Dbp2.

3.4 **Loss of Dbp2 Results in Accumulation of Unspliced Pre-mRNAs**

RNA structures have been reported to impact splicing and alternative splicing in both yeast and mammalian cells (17–23). For example, formation of secondary structures near splice sites of tau pre-mRNAs has been shown to prevent the recognition of the splicing signals by small nuclear RNAs (snRNAs) or small nuclear ribonucleoproteins (snRNPs), thereby inhibiting splicing (24–

26). DEAD-box helicases have been shown to regulate splicing through remodeling of RNA structures (24, 27–31). Specifically, the helicase activity of human DDX5 is required to unwind the stem-loop structure at the 5' splice site (5' ss) of exon X in tau pre-mRNA. As a result, DDX5 facilitates the recognition of 5'ss by U1 snRNA, thereby regulating alternative splicing (24). In *S. cerevisiae*, an earlier study also described a splicing defect in *RPL18* transcript in the absence of *DBP2* (32), consistent with the functional role of its ortholog DDX5 in humans.

To study the role of Dbp2 in splicing, we first asked if Dbp2 associates with intron-containing pre-mRNAs at specific regions, by analyzing the ~250 yeast genes containing one intron as a separate group. This revealed a distinct Dbp2-binding cluster within the second exon near the 3' splice site (Figure 3.3A). Moreover, we found that more than half of the 257 intron-containing mRNAs are Dbp2-binding targets. These observations suggest that Dbp2 may also function in regulation of splicing, similar to its human counterpart DDX5 (30, 33–35).

We then asked if *DBP2* is necessary for efficient splicing. To assess splicing efficiency, we calculated the intron accumulation index for intron-containing genes between wild type and *dbp2Δ* cells as previously described (36). Nearly half (43%) of the intron-containing transcripts in our RNA seq exhibit intron accumulation in the absence of *DBP2* (Figure 3.3B). Importantly, Dbp2-binding targets are significantly enriched in these defective transcripts, indicating a correlation of Dbp2 binding and splicing efficiency (Figure 3.3B, p-value = $3.58e^{-9}$). We also found evidence for improved splicing efficiency in *dbp2Δ* as compared to wild type, however, improved splicing was not well correlated with binding (Figure 3.3B). This improved splicing is reminiscent of the “hungry spliceosome” model (36), whereby reduced splicing of highly expressed pre-mRNAs enables the spliceosome to act more efficiently on less abundant, inefficiently spliced gene products. Consistently, over 56% of the pre-mRNAs exhibiting Dbp2-dependent splicing encode ribosomal proteins (Figure 3.3C-D), indicating that splicing improvement is indirect.

Similar to transcription termination, not all intron-containing pre-mRNAs exhibit splicing defects in the absence of *DBP2*. One of the factors that impact splicing efficiency is intron size (37, 38). Thus, we plotted the distribution of splicing defective transcripts in *dbp2Δ* cells by the length of their respective introns (Figure 3.3E). Strikingly, transcripts with Dbp2-dependent splicing efficiency show a propensity to have long introns (median = 349 nt) compared to all single intron-containing pre-mRNAs (median = 148 nt). This is evident when plotting the quartile distribution of the intron size for intron-containing genes with defects in splicing in the absence of

DBP2 (Figure 3.3F). This suggests that there is an increased requirement for Dbp2 in pre-mRNA processing with long introns.

3.5 Dbp2 Enzymatic Activity is Necessary for Efficient Splicing

To independently validate splicing defects in *dbp2Δ*, we conducted RT-qPCR to detect the ratio of unspliced pre-mRNA to total mRNA of selected transcripts *RPL25* and *RPL30*. This revealed accumulation of reads from the intron in the absence of *DBP2* (Figure 3.4A). Furthermore, we observed rescue of pre-mRNA splicing defects to wild type levels only when expressing a catalytically active Dbp2 and not the ATPase-deficient *dbp2-E268Q* variant (Figure 3.4B). However, we did observe a small, statistically significant decrease in the amount of unspliced *RPL25* with the ATPase-deficient *dbp2-E268Q* variant relative to empty vector alone. As the E268Q mutation is a Motif II mutation that allows ATP binding but inhibits hydrolysis and recycling from RNA (39), this could be due to partial rescue during a single round of duplex unwinding (40).

3.6 Future Plans for Study of Dbp2 in Splicing

3.6.1 To Characterize snRNA Expression in *dbp2Δ* Compared to WT

Small nuclear RNAs (snRNAs) are essential for splicing reactions (41). In our iCLIP data, we found that all the five snRNAs in *S. cerevisiae* are identified as Dbp2 binding targets (Chapter 2). In addition, these RNAs are targeted by the Nrd1-Nab3-Sen1 (NNS) complex for the termination of transcription (16). Therefore, it is likely that Dbp2 functions in the termination and processing of snRNA by coupling with the NNS complex and thereby impacting splicing of pre-mRNAs. To test this possibility, RT-qPCR can be performed to study whether the expression levels of snRNAs are altered in the absence of *DBP2*. Northern blots can also be conducted to visualize the lengths of snRNAs in *dbp2Δ* and wild type to test if the processing of snRNA is dependent on Dbp2.

3.6.2 To Compute the Folding Energy of Dbp2-Bound Introns Comparing to Introns Without Detectable Dbp2 Binding

To test if Dbp2 targeted introns are more likely to form stable secondary structures than introns without Dbp2 binding, computational methods, such as RNAfold in the ViennaRNA package (42), can be utilized to calculate the structure with minimum free energy for each intron

sequence. We were unable to confirm structural rearrangement of pre-mRNAs in *dbp2Δ* cells using our Structure-seq data due to low read coverage across introns in wild type cells as a result of rapid degradation of unprocessed pre-mRNAs (43, 44). However, computational methods have been widely used to uncover the potential biological functions of RNA structures (45, 46), and therefore could also be used to find the common characteristics of Dbp2 targeted introns.

3.6.3 To test Dbp2-dependent splicing of specific genes using the *CUP1* reporter system

The *CUP1* reporter system used for our termination assay has been utilized to determine splicing efficiency previously (47). Therefore, by replacing the *ACT1* exon-intron-exon region with sequences around and including the intron of *RPL25* or *RPL30*, the two tested transcripts with *DBP2*-dependent splicing (Fig. 3.5), we would be able to determine the splicing efficiency in wild type and *dbp2Δ* with growth assay in copper containing media. Both *RPL25* and *RPL30* introns are predicted to form secondary structures using RNAfold (42). Thus, mutations that destabilize the structure can be introduced and in the reporter and test if the splicing defect in *dbp2Δ* can be rescued. This would indicate whether *DBP2*-dependent splicing involves RNA structural remodeling.

3.7 Perspectives

RNAPII pausing at 5' end of genes can be a rate-limiting step in RNA synthesis, impacting mRNA abundance (49). In addition, premature termination at 5' ends is also connected to RNAPII pausing, serving as a mechanism of gene regulation (50, 51). Our observation of *DBP2*-dependent RNAPII pausing and recruitment of Nrd1 at 5' ends indicates that Dbp2 may have a role in regulating RNAPII kinetics through its function in transcription attenuation (Figure 3A). Characterization of the direct link between Dbp2 and RNAPII pausing will provide a new paradigm of how DEAD-box helicases determine gene expression at the initial stage of mRNA metabolism.

Both Dbp2 and its human ortholog DDX5 have been found in pre-catalytic spliceosomes purified by affinity selection (52, 53). Furthermore, DDX5 associates with other splicing regulators (54) and may participate in alternative splicing by remodeling structures at splice sites, as shown in the *tau* and *H-ras* pre-mRNAs (30, 36). Consistently, our data argues that Dbp2 also promotes efficient pre-mRNA splicing (Figure 3.3), and this activity is dependent on the helicase activity

(Figure 3.4B). In addition to splice site remodeling, Dbp2 may remodel the structure of the intron, as studies have shown that secondary structures in introns can bring splicing signals in proximity to promote splicing (23, 55). This would be consistent with the preference for *DBP2*-dependent splicing of long introns (Figure 3.3E). Our study will provide the molecular mechanism of how Dbp2 functions in splicing, revealing a previously unknown factor involved in the processing of intron-containing pre-mRNAs.

3.8 References

1. Gromak N, West S, Proudfoot NJ. 2006. Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II. *Mol Cell Biol* 26:3986–3996.
2. Hyman LE, Moore CL. 1993. Termination and pausing of RNA polymerase II downstream of yeast polyadenylation sites. *Mol Cell Biol* 13:5159–5167.
3. Wagschal A, Rousset E, Basavarajaiah P, Contreras X, Harwig A, Laurent-Chabalier S, Nakamura M, Chen X, Zhang K, Meziane O, Boyer F, Parrinello H, Berkhout B, Terzian C, Benkirane M, Kiernan R. 2012. Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* 150:1147–1157.
4. Kuehner JN, Brow DA. 2008. Regulation of a Eukaryotic Gene by GTP-Dependent Start Site Selection and Transcription Attenuation. *Mol Cell* 31:201–211.
5. Arigo JT, Carroll KL, Ames JM, Corden JL. 2006. Regulation of yeast NRD1 expression by premature transcription termination. *Mol Cell* 21:641–651.
6. Jenks MH, O’Rourke TW, Reines D. 2008. Properties of an Intergenic Terminator and Start Site Switch That Regulate IMD2 Transcription in Yeast. *Mol Cell Biol* 28:3883–3893.
7. Tedeschi FA, Cloutier SC, Tran EJ, Jankowsky E. 2018. The DEAD-box protein Dbp2p is linked to noncoding RNAs, the helicase Sen1p, and R-loops. *RNA* 24:1693–1705.
8. Barrass JD, Reid JEA, Huang Y, Hector RD, Sanguinetti G, Beggs JD, Granneman S. 2015. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biol* 16.
9. Garibaldi A, Carranza F, Hertel KJ. 2017. Isolation of newly transcribed rna using the metabolic label 4-thiouridine, p. 169–176. *In* *Methods in Molecular Biology*.

10. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* 38:576–589.
11. Harlen KM, Churchman LS. 2017. The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol*.
12. Collin P, Jeronimo C, Poitras C, Robert F. 2019. RNA Polymerase II CTD Tyrosine 1 Is Required for Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. *Mol Cell* 73:655–669.e7.
13. Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* 15:2147–2158.
14. Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. 2008. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 15:795–804.
15. Carroll KL, Ghirlando R, Ames JM, Corden JL. 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* 13:361–373.
16. Arndt KM, Reines D. 2015. Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu Rev Biochem* 84:381–404.
17. Mueller N, van Bel N, Berkhout B, Das AT. 2014. HIV-1 splicing at the major splice donor site is restricted by RNA structure. *Virology* 468:609–620.
18. Singh NN, Singh RN, Androphy EJ. 2007. Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res* 35:371–389.
19. Blanchette M, Chabot B. 1997. A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *RNA* 3:405–419.
20. Watakabe A, Inoue K, Sakamoto H, Shimura Y. 1989. A secondary structure at the 3' splice site affects the in vitro splicing reaction of mouse immunoglobulin μ chain pre-mRNAs. *Nucleic Acids Res* 17:8159–8169.
21. Charpentier B, Rosbash M. 1996. Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. *RNA* 2:509–22.

22. Muh SJ, Hovhannisyan RH, Carstens RP. 2002. A non-sequence-specific double-stranded RNA structural element regulates splicing of two mutually exclusive exons of fibroblast growth factor receptor 2 (FGFR2). *J Biol Chem* 277:50143–50154.
23. Krehling JM, Graveley BR. 2005. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila* Dscam pre-mRNA. *Mol Cell Biol* 25:10251–10260.
24. Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, Liu Z, Chen S, Wu JY. 2011. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol* 31:1812–1821.
25. Donahue CP, Muratore C, Wu JY, Kosik KS, Wolfe MS. 2006. Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. *J Biol Chem* 281:23302–23306.
26. Yamashita T, Tomiyama T, Li Q, Numata H, Mori H. 2005. Regulation of tau exon 10 splicing by a double stem-loop structure in mouse intron 10. *FEBS Lett* 579:241–244.
27. Staley JP, Guthrie C. 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* 3:55–64.
28. Honig A, Auboeuf D, Parker MM, O'Malley BW, Berget SM. 2002. Regulation of Alternative Splicing by the ATP-Dependent DEAD-Box RNA Helicase p72. *Mol Cell Biol* 22:5698–5707.
29. Lin C, Yang L, Yang JJ, Huang Y, Liu ZR. 2005. ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol Cell Biol* 25:7484–7493.
30. Camats M, Guil S, Kokolo M, Bach-Elias M. 2008. P68 RNA helicase (DDX5) alters activity of Cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS One* 3.
31. Guil S, Gattoni R, Carrascal M, Abián J, Stévenin J, Bach-Elias M. 2003. Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol Cell Biol* 23:2927–2941.
32. Barta I, Iggo R. 1995. Autoregulation of expression of the yeast Dbp2p “DEAD-box” protein is mediated by sequences in the conserved DBP2 intron. *EMBO J* 14:3800–3808.
33. Camats M, Guil S, Kokolo M, Bach-Elias M. 2008. P68 RNA helicase (DDXS) alters activity of Cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS One* 3.

34. Dardenne E, PolayEspinoza M, Fattet L, Germann S, Lambert MP, Neil H, Zonta E, Mortada H, Gratadou L, Deygas M, Chakrama F, Samaan S, Desmet FO, Tranchevent LC, Dutertre M, Rimokh R, Bourgeois CF, Auboeuf D. 2014. RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Rep* 7:1900–1913.
35. Zonta E, Bittencourt D, Samaan S, Germann S, Dutertre M, Auboeuf D. 2013. The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. *Nucleic Acids Res* 41:554–564.
36. Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, Liu Z, Chen S, Wu JY. 2011. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol* 31:1812–1821.
37. Munding EM, Shiue L, Katzman S, Donohue J, Ares M. 2013. Competition between Pre-mRNAs for the splicing machinery drives global regulation of splicing. *Mol Cell* 51:338–348.
38. Klinz FJ, Gallwitz D. 1985. Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res* 13:3791–3804.
39. Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin E V., Kondrashov FA. 2002. Selection for short introns in highly expressed genes. *Nat Genet*.
40. Cloutier SC, Ma WK, Nguyen LT, Tran EJ. 2012. The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. *J Biol Chem* 287:26155–26166.
41. Liu F, Putnam A, Jankowsky E. 2008. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A* 105:20209–20214.
42. Valadkhan S. 2005. snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol*.
43. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* 6:26.
44. Gudipati RK, Xu Z, Lebreton A, Séraphin B, Steinmetz LM, Jacquier A, Libri D. 2012. Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. *Mol Cell* 48:409–421.
45. Sayani S, Janis M, Lee CY, Toesca I, Chanfreau GF. 2008. Widespread Impact of Nonsense-Mediated mRNA Decay on the Yeast Intronome. *Mol Cell* 31:360–370.

46. Fallmann J, Will S, Engelhardt J, Grüning B, Backofen R, Stadler PF. 2017. Recent advances in RNA folding. *J Biotechnol* 261:97–104.
47. Schroeder SJ. 2018. Challenges and approaches to predicting RNA with multiple functional structures. *RNA*.
48. Lesser CF, Guthrie C. 1993. Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1. *Genetics* 133:851–863.
49. Mayer A, Landry HM, Churchman LS. 2017. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr Opin Cell Biol*.
50. Henriques T, Gilchrist DA, Nechaev S, Bern M, Muse GW, Burkholder A, Fargo DC, Adelman K. 2013. Stable pausing by rna polymerase II provides an opportunity to target and integrate regulatory signals. *Mol Cell* 52:517–528.
51. Natarajan M, Schiralli Lester GM, Lee C, Missra A, Wasserman GA, Steffen M, Gilmour DS, Henderson AJ. 2013. Negative elongation factor (NELF) coordinates RNA polymerase II pausing, premature termination, and chromatin remodeling to regulate HIV transcription. *J Biol Chem* 288:25995–26003.
52. Warkocki Z, Odenwälder P, Schmitzová J, Platzmann F, Stark H, Urlaub H, Ficner R, Fabrizio P, Lührmann R. 2009. Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat Struct Mol Biol* 16:1237–1243.
53. Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, Urlaub H, Stark H, Lührmann R. 2010. Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* 16:2384–403.
54. Damianov A, Ying Y, Lin CH, Lee JA, Tran D, Vashisht AA, Bahrami-Samani E, Xing Y, Martin KC, Wohlschlegel JA, Black DL. 2016. Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell* 165:606–619.
55. Gahura O, Hammann C, Valentová A, Puta F, Folk P. 2011. Secondary structure is required for 3' splice site recognition in yeast. *Nucleic Acids Res* 39:9759–9767.

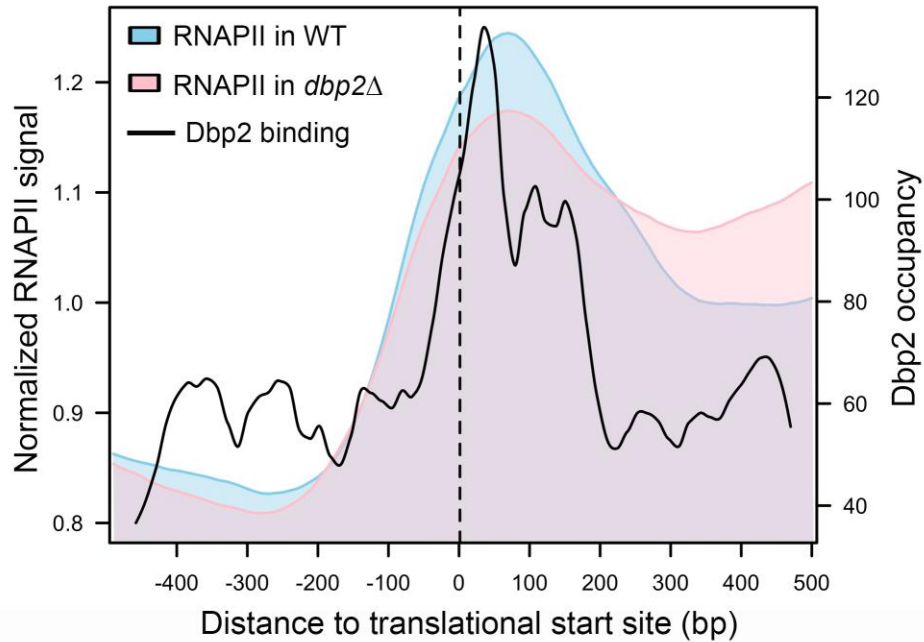


Figure 3.1. Dbp2 binding correlates with RNAPII pausing at 5' ends of protein-coding genes.

Normalized RNAPII occupancy across translational start sites of mRNAs from RNAPII ChIP-seq in the wild type (blue) and *dbp2Δ* (pink). Coordinates of translational start sites were obtained from SGD. Distribution of Dbp2 binding sites (black line) was derived from Dbp2 iCLIP-seq data (Chapter 2).

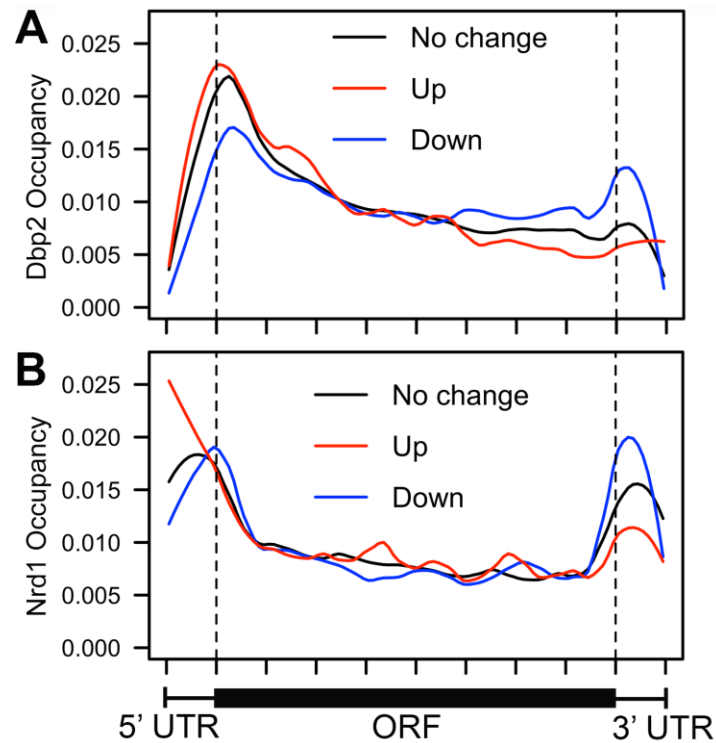


Figure 3.2. Dbp2 and Nrd1 preferentially binds to the 5' ends of mRNAs up-regulated in *dbp2Δ*, and display diminished 3' binding compared to the other mRNAs.

Meta-analysis of Dbp2 (A) and Nrd1 (B) in mRNAs up-regulated (red), down-regulated (blue), without change in expression (black) in *dbp2Δ* as compared to the wild type. Dashed vertical lines mark boundaries of the coding region.

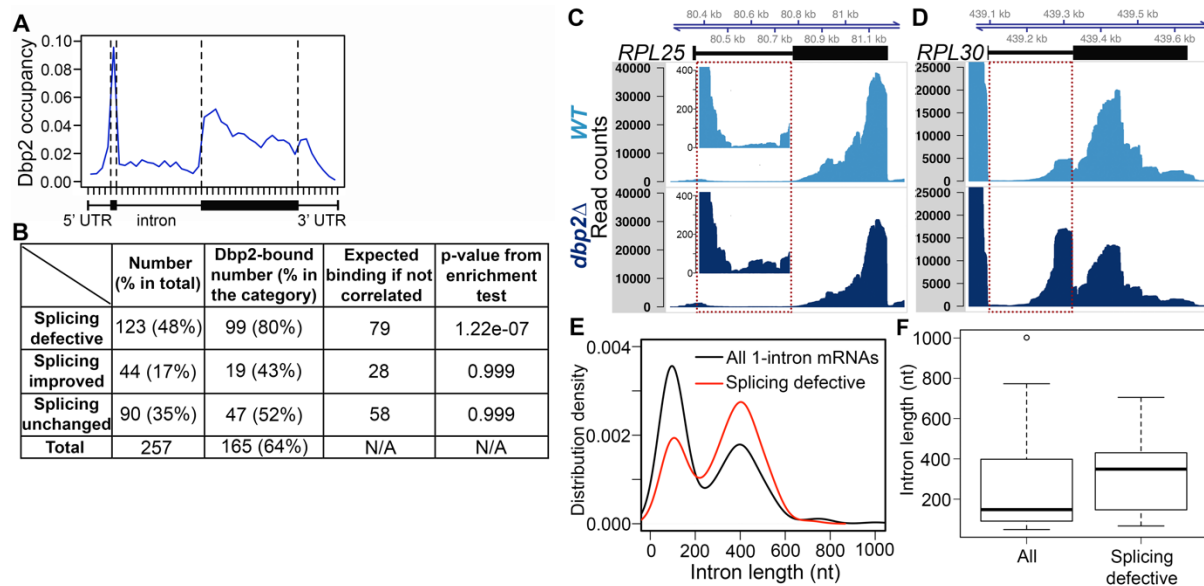


Figure 3.3. Dbp2 promotes efficient pre-mRNA splicing.

(A) Meta-analysis of Dbp2 crosslinking sites across single intron-containing transcripts. (94% of intron-containing genes only have one intron.) Dashed vertical lines represent the start of the first exon, the 5' splice site, the 3' splice site, and the start of the 3' UTR. (B) Analysis of the change in splicing efficiency in the absence of *DBP2* compared to wild type. For each intron-containing mRNA, the intron accumulation index (IAI) was calculated by comparing the read counts in the intron and exon as previously described (20) and grouped into defective, improved, and unchanged in the absence of *DBP2*. The expected number of Dbp2 binding transcripts for the null hypothesis was calculated using the `chisq.test` function in R. A Fisher's one-sided exact test was performed to test the enrichment of Dbp2 binding targets in each group (p-value, rightmost column). (C-D) Genomic tracks of aligned reads from intron-containing genes, *RPL25* and *RPL30*, from RNA seq of *dbp2Δ* and wild type cells. Red boxes highlight regions with accumulated reads in the absence of DBP2. In B, the read coverage in the intron of *RPL25* is magnified (inset) for visualization. (D) The distribution of intron lengths in all intron-containing genes (black) and mRNAs with a splicing defect (red) in *dbp2Δ*. The shape of the distribution was produced using the Kernel density plot in R. (E) The quartile distribution of intron lengths as box plots. For (D) and (E), only transcripts with one intron are shown for simplicity.

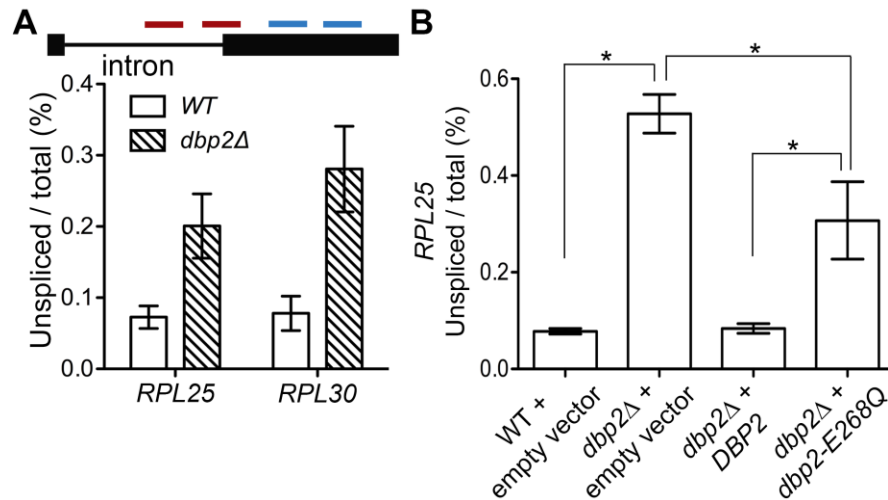


Figure 3.4. The enzymatic activity of Dbp2 is required for efficient pre-mRNA splicing.

(A) RT-qPCR analysis of the ratio of unspliced pre-mRNA versus total in wild type and *dbp2Δ* cells. Two primer sets, one in the intron (red) and the other in the second exon (blue), were used to detect the unspliced and total mRNA, respectively (schematic). (B) RT-qPCR of *RPL25* unspliced pre-mRNA/mRNA ratio of in wild type or *dbp2Δ* strains transformed with empty vector, or plasmids expressing wild-type *DBP2* or an ATPase-deficient (E268Q) *dbp2* allele. Values correspond to the average from three biological replicates \pm one SD. Asterisks indicate a one- sided p-value of < 0.05 .

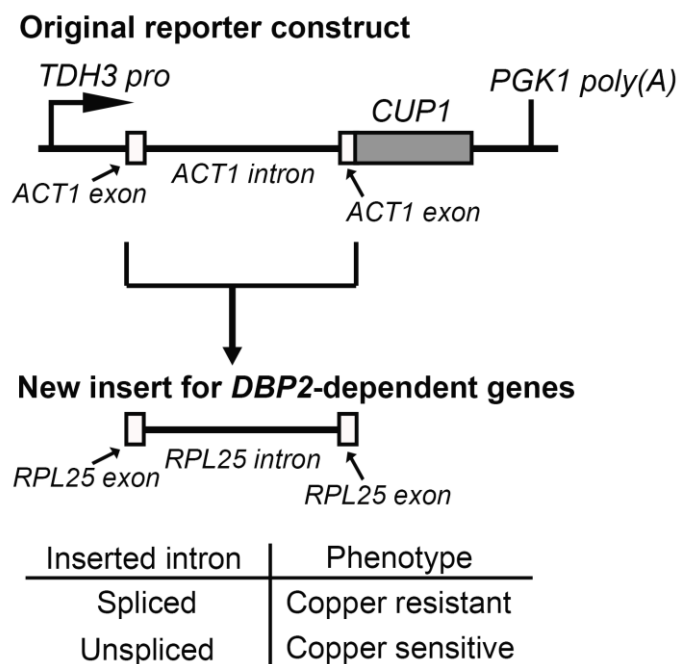


Figure 3.5. Schematic representation of the splicing reporter.

Sequences encompass the intron region of genes with *DBP2*-dependent splicing efficiency can be cloned to replace the *ACT1* sequence in the original construct (30). The unspliced transcripts would be unstable and degraded. Thus, the splicing efficiency would be correlated with the level of copper resistance of cells.

DECLARATION OF COLLABORATIVE WORK

Chapter 1. Introduction

I wrote this chapter. Dr. Elizabeth Tran and Dr. Pete Pascuzzi provided comments for this chapter.

Chapter 2. Genome-Wide Discovery of DEAD-Box RNA Helicase Targets Reveals RNA Structural Remodeling in Transcription Termination

Chapter 2 is a version of the research article published in the Genetics journal entitled “Genome-Wide Discovery of DEAD-Box RNA Helicase Targets Reveals RNA Structural Remodeling in Transcription Termination”. This paper is authored by Yu-Hsuan Lai, Krishna Choudhary, Sara C. Cloutier, Zheng Xing, Sharon Aviran and Elizabeth J. Tran. This article is distributed with the permission from Genetics Society of America.

Chapter 3. Unpublished results

I produced data for this section and wrote this section. Dr. Elizabeth Tran provided extensive advice.

VITA

Education

PhD in Biochemistry, 2014- 2019

Purdue University, West Lafayette, IN

MS in Biochemical Science and Technology, 2011- 2013

National Taiwan University, Taipei, Taiwan

BS in Biochemical Science and Technology, 2007- 2011

National Taiwan University, Taipei, Taiwan

Professional Experiences

Research Assistant, Purdue University, West Lafayette, IN, Aug. 2014 – May 2019

Teaching Assistant, Purdue University, West Lafayette, IN, Jan. 2016 – May 2019

Research Assistant, National Taiwan University, Taipei, Taiwan, Jan. – July 2014

Research Assistant, Tunghai University, Taichung, Taiwan, Sept. – Dec. 2013

Undergraduate Research, National Taiwan University, Taipei, Taiwan, Aug. 2009-May 2011

Publications

1. **Yu-Hsuan Lai**, Sara Cloutier, Zheng Xing, Krishna Choudhary, Sharon Aviran*, Elizabeth Tran*. *Genome-wide Discovery of DEAD-box RNA Helicase Targets Reveals RNA Structural Remodeling in Transcription Termination*. Genetics, 2019, DOI: 10.1534/genetics.119.302058.
2. Krishna Choudhary, **Yu-Hsuan Lai**, Elizabeth J. Tran and Sharon Aviran*. *dStruct: identifying differentially reactive regions from RNA structurome profiling data*. Genome Biology, 2019, DOI: 10.1186/s13059-019-1641-3.
3. Chun-Ching Lee, Yu-Chieh Liao, **Yu-Hsuan Lai**, and Min-Chieh Chuang*. *Recognition of Dual Targets by a Molecular Beacon-Based Sensor: Subtyping of Influenza A Virus*. Analytical Chemistry, 2015, DOI: 10.1021/acs.analchem.5b00810
4. **Yu-Hsuan Lai**, Sin-Cih Sun and Min-Chieh Chuang*. *Biosensors with built-in biomolecular logic gates for practical applications*. Biosensors (Basel), 2014, DOI: 10.3390/bios4030273.

5. **Yu-Hsuan Lai**, Chang-Chun Lee, Chwan-Chuen King, Min-Chieh Chuang* and Ja-an Annie Ho*. *Exploitation of stem-loop folded DNA toward dual-input genetic analysis: Extension to subtyping of influenza viruses*. Chemical Science, 2014, DOI: 10.1039/c4sc01289e.
6. **Yu-Hsuan Lai**, Chieh-Hua Lin, Yu-Chieh Liao, Tsui-Ming Kuo and Min-Chieh Chuang*. *Biomolecular logic gate for analysis of the New Delhi metallo- β -lactamase (NDM)-coding gene with concurrent determination of its drug resistance-encoding fragments*. Chemical Communication, 2014, DOI: 10.1039/C4CC01108B.
7. Ja-an Annie Ho*, **Yu-Hsuan Lai**, Li-Chen Wu, Shen-Huan Liang, Song- Ling, Wong and Jr-Jiun Liou, 2012. *Analysis of Biotin (Vitamin B7) and Folic Acid (Vitamin B9): A Focus on Immunosensor Development with Liposomal Amplification*. Food and Nutritional Components in Focus, Royal Society of Chemistry Publishing.

Conference Presentations

-
- 2018 **Yu-Hsuan Lai**, Krishna Choudhary, Sharon Aviran*, Elizabeth J. Tran*. *Genome-wide Identification of DEAD-box RNA Helicase Targets Reveals Roles for RNA Secondary Structure Remodeling in transcription termination*. Poster presentation at the RNA society annual meeting, Berkeley, CA.
- 2017 **Yu-Hsuan Lai**, Krishna Choudhary, Sharon Aviran*, Elizabeth J. Tran*. *Genome-wide Identification of DEAD-box RNA Helicase Targets Reveals Roles for RNA Secondary Structure Remodeling in mRNA Processing*. Poster presentation at the Rustbelt RNA meeting, Indianapolis, IN.
- 2017 **Yu-Hsuan Lai**, Krishna Choudhary, Sharon Aviran*, Elizabeth J. Tran*. *Genome-wide Identification of DEAD-box RNA Helicase Targets Reveals Roles for RNA Secondary Structure Remodeling in mRNA Processing*. Oral presentation at the Cold Spring Harbor meeting on Eukaryotic mRNA Processing, Cold Spring Harbor, NY.
- 2016 **Yu-Hsuan Lai**, Siwen Wang, Elizabeth J. Tran*. *Characterizing the role of the DEAD-box protein Dbp2 in RNA 3' end processing*. Poster presentation at the Rustbelt RNA Meeting, Cleveland, OH.
- 2015 **Yu-Hsuan Lai**, Elizabeth J. Tran*. *The role of the DEAD-box protein Dbp2 in RNA structure remodeling*. Poster presentation at the Rustbelt RNA Meeting, Huron, OH

2013 **Yu-Hsuan Lai**, Min-Chieh Chuang* and Ja-an Annie Ho*. *Molecular Processing in Biomedical Diagnostics: A Bioelectrochemical Genosensor Constructed with Enzyme Logic Gate for Digital Subtyping of Specific Influenza Virus*. Poster presentation at the International Conference on Life Science and Biological Engineering, Tokyo, Japan

2013 **Yu-Hsuan Lai**, Chang-Chun Lee, Chwan-Chuen King, Min-Chieh Chuang* and Ja-an Annie Ho*. *Exploiting structurally controllable stem-loop folded DNA toward dual-input genetic analysis: Extension to subtyping of influenza viruses*. Poster presentation at the Annual Meeting of Chemical Society, Taipei, Taiwan

Honors and Awards

2017-2018	Graduate Research Training Assistantship for Bioinformatics, Center for Cancer Research, Purdue University
2017	Summer research grant, College of Agriculture, Purdue University
2017-2018	Bird Stair Graduate Research Fellowship, Department of Biochemistry, Purdue University
2017	RNA society travel fellowship
2013	Dean's Award of College of Life Science, National Taiwan University
2013	Excellent Scientific Poster Award of College of Life Science, National Taiwan University
2010-2011	College Student Research Scholarship, National Science Council (NSC), Taiwan
2010-2011	Presidential Award, Department of Biochemical Science and Technology, National Taiwan University

Service

2017-2018	Professional Development Chair, Graduate Student Organization in the Department of Biochemistry, Purdue University
2017	Annual Retreat Committee for the Department of Biochemistry, Purdue University
2015-2016	Vice President, Taiwanese graduate student association at Purdue University
2016	Student invited seminar committee in Department of Biochemistry, Purdue University
2009-2010	President, student association in department of Biochemical Science and Technology, National Taiwan University