INFLUENCE OF RNA POLYMERASE II CATALYTIC ACTIVITY ON TRANSCRIPTION START SITE SELECTION

A Dissertation

by

HUIYAN JIN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Craig D. Kaplan
Committee members,	David O. Peterson
	Paul D. Straight
	Matthew S. Sachs
Head of department,	Gregory D. Reinhart

August 2015

Major Subject: Biochemistry

ABSTRACT

RNA Polymerase II (Pol II) is responsible for expression of all protein-coding genes in eukaryotes. To understand gene expression at the molecular level, it is essential to understand the mechanisms of Pol II function. I investigate how Pol II catalytic activity influences the first step of gene expression, transcription initiation, in *Saccharomyces cerevisiae.* My dissertation focuses on the mechanisms by which Pol II activity defects contribute to transcription start site (TSS) selection/utilization and promoter output. I utilize mutants with substitutions in the Pol II active site that have different elongation rates *in vitro* and show varied growth phenotypes *in vivo* that correlate with the observed alterations in elongation rates. I employ genetic and biochemical approaches to investigate the relationships between the TSS phenotypes of Pol II activity/General Transcription Factor (GTF) mutants and conditional and general growth phenotypes. I show that while some conditional growth phenotypes correlate closely with TSS defects, TSS defects are not likely the main determinant for general growth defects. I further explore growth phenotypes and TSS defects between Pol II genetic interactors and GTF mutants combined with Pol II activity mutants and present different models for the relationships discovered so far. I discover a novel function of Sub1, which is suggested to play positive roles in transcription as an initiation factor, altering TSS utilization on its own and modifying TSS defects conferred by Pol II activity mutants. I take a genome wide approach to map TSSs, general transcription factor occupancies, and nucleosome positions to investigate the mechanism of Pol II

activity control over initiation, and determine how promoter architecture influences TSS selection/utilization and nucleosome positioning. I show that promoter architecture is a critical parameter in determination of various initiation properties including transcription output (gene expression) levels.

ACKNOWLEDGEMENTS

I thank my advisor Dr. Kaplan for all the teaching, guidance, support and patience in past five years I have been working with him. I learnt not only how to be a researcher but also how to be a better version of myself and I am truly grateful for all the time he invested helping me grow.

I thank my committee members Dr. Sachs, Dr. Peterson and Dr. Straight and previous committee member Dr. Feng Qiao for their helpful suggestions and guidance on developing this dissertation. I thank Dr. Schwartz (University of Texas) for taking out his time on helping me on some bioinformatics problems. I thank Kaplan lab members for their assistance in some experiments and helpful discussions. I was glad to be part of Kaplan lab. I thank our collaborator Dr. Nickels' lab (Rutgers University) for TSS sequencing and Dr. Pugh's lab (Penn State University) for ChIP-exo sequencing. I also thank Jordi Abante (TAMU) for bioinformatics analysis for nucleosome sequencing. I thank Dr. Haberle (Imperial College London) for generously providing raw data for zebrafish TSS sequencing for my analyses.

I thank my mother and father for all the love and motivation. Last but not least, I thank all the close friends for their valuable friendship and encouragement to make graduate school easier.

NOMENCLATURE

Pol II	RNA Polymerase II
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
TSS	Transcription Start Site
GTF	General Transcription Factor
TFII	Transcription Factor II
PIC	Pre-Initiation Complex
TBP	TATA Binding Protein
TAF	TBP-associated factor
CPE	Core Promoter Element
CP	Core Promoter
MPA	Mycophenolic Acid
TL	Trigger Loop
Y	Pyrimidine
R	Purine
nt	Nucleotide
cDNA	complementary DNA
A	Adenine
T	Thymine
G	Guanine

C	Cytosine
SOLiD	Sequence by Oligonucleotide Ligation and Detection
ChIP-exo	Chromatin Imunoprecipitation-exonuclease

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

Page

Page

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Transcription is the process of creating an RNA copy (a transcript) of a DNA template. Multi-subunit RNA Polymerases (msRNAPs) are keys to transcription machineries in eukaryotes, with RNA Polymerase II (Pol II) being responsible for transcription of protein coding genes (mRNA) and varied types of capped non-coding RNA. Regulation of Pol II transcription is the first step in gene expression and also a critical component of cellular regulation (Cramer et al., 2001). The transcription process includes several steps: assembly of a Pre-Initiation Complex (PIC), melting of promoter DNA, initiation, elongation and termination (Freire-Picos et al., 2005). All these processes are tightly regulated and are carried out with the contributions of many factors in a complex harmony. Thus understanding how Pol II functions and its regulation along with cotranscriptional events in various stages of transcription has been a key task for study of gene expression. Because msRNAPs in different organisms are highly similar in their core structural scaffolds, having mechanistic insights into *Saccharomyces cerevisiae* Pol II sheds valuable light on other msRNAPs. The Kaplan lab is interested in the influence of Pol II catalytic activity on gene expression *in vivo* using the model organism *S. cerevisiae*. My dissertation mainly focuses on understanding the influence of Pol II catalytic activity on transcription start site (TSS) selection and aspects of Pol II initiation regulation. I first describe how Pol II mutants are used for alteration of Pol II activity and how they are used to investigate the influence of Pol II activity on gene

expression *in vivo*. I then describe our current understanding of Pol II initiation and TSS selection in *S. cerevisiae* and the open questions my dissertation endeavors to explore.

Control of Pol II catalytic activity and substrate selection through trigger loop

The key enzyme in eukaryotic transcription machinery, Pol II, is a large protein complex (~500 kDa) with 12 subunits, with Rpb1 being the largest (~200 kDa) essential subunit. Catalytic activity of Pol II is controlled by a conserved mobile domain in Rpb1 named Trigger loop (TL). The TL's primary functions include substrate selection, transcription fidelity, nucleotide catalysis, translocation and intrinsic cleavage (Kaplan et al., 2008; Kireeva et al., 2008; Larson et al., 2012; Nayak et al., 2013; Toulokhonov et al., 2007; Zhang et al., 2010). During transcription elongation, which consists of repeated nucleotide addition cycles, the TL "closes" towards Pol II active site, possibly induced by matched NTP substrate binding, trapping the NTP substrate within the Pol II active site. TL-substrate interactions promote the catalysis of phosphodiester bond formation. Subsequently, the TL transitions to an "open" conformation allowing and possibly promoting PPi release. This "open" conformation has been proposed to allow translocation of Pol II to the next template position for repetition of the nucleotide addition cycle (Da et al., 2012; Malinen et al., 2012). Transition of these dynamic states requires interactions among TL residues and between the TL and other Pol II residues/domains (Silva et al., 2014; Wang et al., 2013), suggesting a complex functional network of interactions. Previous studies have discovered alterations of residues in TL result in changes in catalytic activity and fidelity of the enzyme (Kaplan et al., 2008;

Kireeva et al., 2008; Tan et al., 2008). The Kaplan lab has investigated a large number of substitutions in the TL and adjacent residues for biochemical and genetic phenotypes and found a set of TL mutants that alters catalytic activity *in vitro* (Kaplan et al., 2012; Kaplan et al., 2008). This set of Pol II mutants is a valuable tool for exploration of Pol II activity control of gene expression, and my dissertation utilizes these mutants to focus on the influence of Pol II activity on TSS selection/utilization *in vivo*.

One class of Pol II TL mutants confers faster elongation rates (termed gain of funtion, GOF or "superactive") *in vitro,* while another class confers slower elongation rates (termed loss of function, LOF). The two classes are generally mutually suppressive when combined in genetic interaction studies. Different TL mutants show different growth rates *in vivo*. Interestingly, mutants with catalytic activity closer to WT grow more similarly to WT cells; those with greater catalytic defects show more severe growth defects. Since the magnitude of elongation rate defects correlates well with the extent of defects in cell growth, it suggests that proper control of Pol II catalytic activity is critical for cellular functions. Consistent with this idea, we also obtained Pol II mutants outside of the TL domain that are genetically similar to TL mutants in various genetic interactions. These mutants show a similar correlation between their genetically estimated Pol II activity alteration and their growth phenotypes *in vivo*. The question that arose from discovery of this close correlation of catalytic activity *in vitro* and growth phenotype *in vivo* was what transcriptional defects are causative of these growth defects (Chapter II, III)? Besides altered growth on rich or synthetic media, TL mutants also show various conditional growth phenotypes observable due to the presence of reporter

alleles in our strain. These alleles allow us to examine transcription defects or different responses to cellular transcription substrate depletion by an inducible gene. These conditional growth phenotypes allow us to distinguish between different classes of Pol II mutants. These phenotypes are discussed more in Chapter II, III and (Kaplan et al., 2012). Altering Pol II catalytic activity may have a broad impact in various steps in gene expression and its regulation, and each aspect of gene expression can be probed using a phenotype that mainly originates from that particular aspect. Indeed, Pol II activity mutants conferred various phenotypes both *in vitro* and *in vivo*, including altered TSS selection, RNA splicing, presumed termination or processing defects, and chromosome segregation defects (Braberg et al., 2013; Kaplan et al., 2012; Kaplan et al., 2008; Kireeva et al., 2008; Larson et al., 2012; Viktorovskaya et al., 2013). The severity of TSS selection defects in vivo in both fast and slow Pol II activity mutants correlates well with the extent of their deviation from WT activity. This will be discussed more in the later section.

Pol II activity mutants that have more severely altered activity *in vitro* (both fast and slow) show greater growth defects, more genetic interactions, and greater alterations in gene expression profiles *in vivo* (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012). Growth defects can be suppressed when Pol II fast and slow mutations are combined within the same enzyme; similarly, there is mutual suppression of Pol II mutant TSS distribution defects at *ADH1* in double mutants, indicating a correlation between TSS defects and general growth defects (Kaplan et al., 2012). TL mutants also show wide ranges of genetic interactions with mutants in many transcription factors and

are explored in Chapter II and III. We explored the causative relationship between TSS defects and cell growth and other conditional growth phenotypes in Chapter III.

Pol II initiation and different promoter classes in *S. cerevisiae*

Initiation, the first step in transcription, is highly conserved process with complex regulation requiring many factors (Cramer et al., 2008; Hahn, 2004). Classical biochemical experiments using model promoters have shown that Pol II requires general transcription factors (GTFs)- Transcription Factor II (TFII)D, TFIIB, TFIIF, TFIIE, TFIIH to form a Pre-Initiation Complex (PIC) during transcription initiation. These factors are critical to guide Pol II to a functional core promoter element (CPE) and cooperatively determine efficiency of any particular promoter (Khaperskyy et al., 2008; Wu et al., 1999). Of the GTFs mentioned above, the TATA-binding protein (TBP) subunit of TFIID, TFIIB and TFIIF have been shown to be the minimally required factors for Pol II recruitment to promoter DNA (Flores et al., 1991; Killeen et al., 1992; Pan and Greenblatt, 1994) (Calf thymus Pol II) and for supporting synthesis of runoff transcripts by Pol II when negatively supercoiled template DNAs are used (Parvin and Sharp, 1993; Timmers, 1994; Tyree et al., 1993) (human, Drosophila Pol II). GTFs TFIIH and TFIIE are required for promoter melting, a process that describes unwinding of template dsDNA allowing the ssDNA template (non-coding) strand to be recognized by Pol II. The promoter DNA is bent 90˚ relative to its conformation in closed complex and melted DNA is inserted into the Pol II active site due to the actions of GTFs. Promoter DNA bending occurs upon TBP binding. Interestingly, TBP, a required GTF

for TATA-containing genes, is also in fact required for transcription of all promoters transcribed by Pol II and also for Pol I and Pol III transcription, despite its name (Hernandez, 1993; Sadowski et al., 1993; White and Jackson, 1992).

The 5′ end of the transcript is determined by a process named transcription start site (TSS) selection. Recognition and recruitment of GTFs and Pol II to a core promoter element is presumably a major factor in TSS selection and the other events that result in transcription output. Promoters in *S. cerevisiae* and other eukaryotes can be classified into those that are TATA-containing and those that are TATA-less. TATA-containing promoters contain the most well-known core promoter element, the TATA box (consensus sequence motif TATAWAWR in *S. cerevisiae*, W=A/T, R=A/G), which is highly conserved throughout evolution yet is only found at a subset of promoters (17% in *S. cerevisiae*). TATA-less promoters do not have a consensus TATA motif. Our understanding of the relationships between core promoter elements and TSSs is not complete due to lack of knowledge of core promoter elements or architecture for TATA-less promoters. It is worth noting that many studies of Pol II initiation have focused on TATA-containing promoters. TATA-containing and TATA-less promoters have fundamental differences in their architectures, including sequence compositions of promoter regions, nucleosome positioning, and their dependencies on initiation factors (**Table 1-1**). First, TATA-containing promoters are thymidine (T) rich over the coding strand of the promoter region between TATA element and TSSs. T-richness also correlates with gene expression levels ((Lubliner et al., 2013), Chapter IV), while TATA-less promoters show asymmetric A/T richness lacking the peaked T enrichment

found in TATA-containing promoters (Chapter IV, (Wu and Li, 2010). Second, nucleosome positioning also shows very distinct variation between promoter classes (Kaplan et al., 2009; Rhee and Pugh, 2012; Tirosh and Barkai, 2008; Wu and Li, 2010). The nucleosome is the basic unit of chromatin, the protein/nucleic acid associations that package and compact genomes. The nucleosome consists of double-stranded DNA wrapping around a core of histone proteins. Nucleosomes can be an obstacle for Pol II access to DNA during transcription. For TATA-less promoters, the +1 nucleosome (the first nucleosome downstream of promoter or proximal to the TSS) appears consistently and stably positioned in cells, with a clear NFR (nucleosome-free region)/NDR (nucleosome-depleted region) upstream. The NFR/NDRs for TATA-less promoters are stereotypically positioned with clear edges and are proposed to promote transcription by depleting nucleosome barrier to core promoter access. TATA-containing promoters on the other hand, show a lack of stereotypical +1 nucleosome positioning and no sharp NFR/NDR. Third, these two promoter classes also show differences in their dependence on initiation factor TBP-associated factor 1 (Taf1, a subunit of TFIID); most TATA-less promoters have greater apparent Taf1-dependence and are enriched for its binding, while the majority of TATA-containing promoters are depleted for TAFs (Apone et al., 1998; Green, 2000; Holstege et al., 1998; Li et al., 2000; Rhee and Pugh, 2012). Because of this, TATA-containing and TATA-less promoters can be loosely classified as Taf1-depleted and Taf1-enriched promoter classes respectively.

Table 1-1 Characteristics of two major promoter classes in *S. cerevisiae*

Promoter class	Core promoter element	Dependent factor	Sequence skew	+1 Nucleosome
TATA-containing	TATAWAWR	SAGA complex	AT rich	No peak, not stereotypical
TATA-less	TATA-like element	TBP-associated factor (TAF)s	Asymmetric AT rich	Stereotypical, sharp edged

Transcription start site selection in *S. cerevisiae*

TSS selection at TATA-containing promoters in *S. cerevisiae* involves recognition of start sites positioned 40-120 nucleotides (nt) downstream from the TATA box (if present) and the use of multiple start sites at most promoters ((Basehoar et al., 2004; Corden, 2008; Dvir, 2002; Kang et al., 2015; Li et al., 1994; Malabat et al., 2015; Park et al., 2014; Pelechano et al., 2013; Rojas-Duran and Gilbert, 2012; Struhl, 1987) and Chapter IV) (**Figure 1-1**). Such extensive downstream positioning of TSSs in yeast is distinct from other eukaryotes for TATA-dependent promoters, where start sites for TATA-containing promoters are present in a tightly focused (narrow) window ~30-40 nt downstream of the beginning of the TATA box (with most transcription exactly at +30). However, recent genome wide studies have revealed characteristics of TSSs in other eukaryotes that are similar with those of *S. cerevisiae*. The majority of promoters in metazoans (e.g. *Drosophila melanogaster* (Drosophila), *Caenorhabditis elegans* (*C.*

elegans), zebrafish and human) utilize multiple dispersed TSSs in a manner at least superficially analogous to *S. cerevisiae* (Chen et al., 2013; Consortium et al., 2014; Haberle et al., 2014; Hoskins et al., 2011). In addition, a very recent publication on detection of genome wide TSSs in another model yeast organism, *Schizosaccharomyces pombe* (*S. pombe*), showed utilization of multiple dispersed TSSs similar to *S. cerevisiae* for ~50% promoters (Li et al., 2015). TSSs of *S. pombe* are at a similar distance from TATA elements as in higher eukaryotes for TATA-containing promoters. Despite this difference in TSS distance to promoter element for TATA-containing promoters in *S. cerevisiae*, promoter melting appears to start 20-30 nt downstream of the TATA, similar to higher eukaryotes (Giardina and Lis, 1993; Lada et al., 2013; Pal et al., 2005; Taylor et al., 2013). The initial observation by Giardina and Lis led to proposal that *S. cerevisiae* Pol II scans for favorable TSSs subsequent to promoter melting and open complex formation (see expanded discussion on scanning mechanism below and Chapter V) (**Figure 1-1**). Whether Pol II in other eukaryotes scans for TSSs is unknown (certainly it need not scan when a single TSS at +30 is used). TSSs in other eukaryotes mentioned above can be classified depending on their "shapes" of TSS distribution: sharp/single peaked, focused/unspecified and broad/wide clustered. The classes of focused and broad TSSs are fairly similar to *S. cerevisiae* TSSs (**Figure 1-2**). We hypothesize that TSSs for focused and broad promoters may be selected/utilized by a Pol II scanning mechanism in all eukaryotes (Chapter IV).

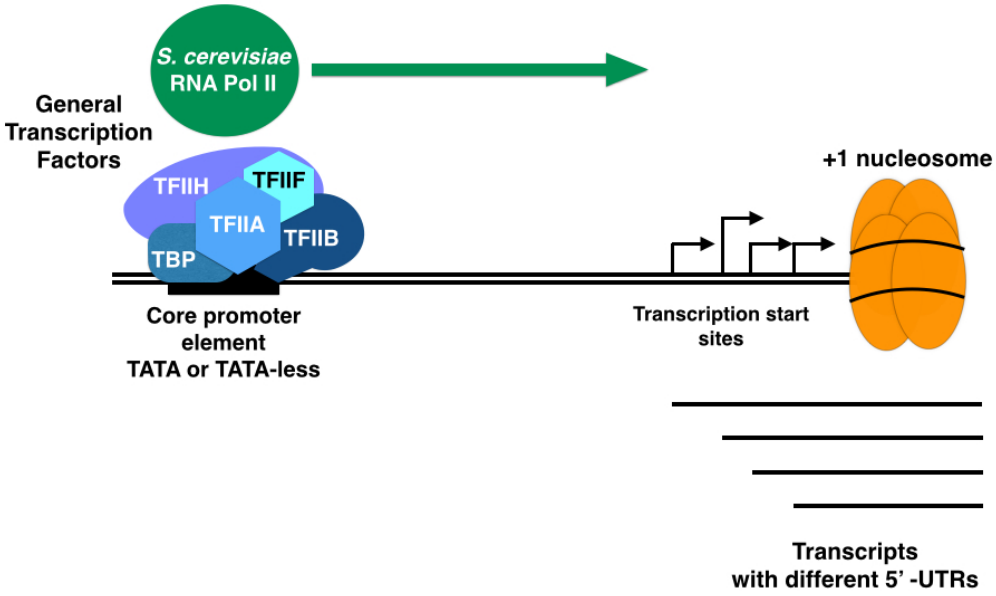


Figure 1-1. Transcription initiation and start site selection in *S. cerevisiae.*

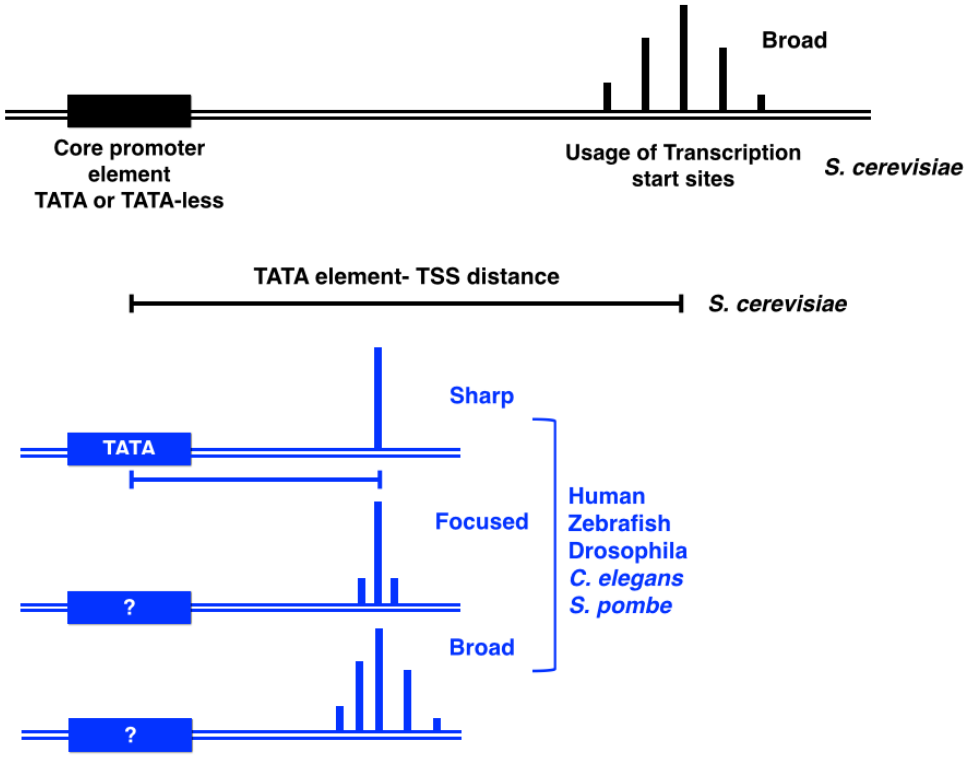


Figure 1-2. Transcription start site selection in *S. cerevisiae* and other eukaryotes.

Proposed core promoter elements in TATA-less promoters

What the functional core promoter elements are for TATA-less promoters still remains unclear. A recent study by the Pugh group proposed that most TATA-less promoters contain a "TATA-like" element that has 1-3 mismatches from the consensus TATA motif TATAWAWR (W=A/T, R=A/G) (Rhee and Pugh, 2012). The authors used a higher resolution ChIP (chromatin imunoprecipitation)-sequencing approach named ChIP-exo (Rhee and Pugh, 2011), utilizing exonucleases to obtain improved resolution for protein localization genome wide (see method in Chapter IV for more details). The authors mapped GTF subunits that are presumably localized to core promoter elements and searched for TATA-like motifs within the GTF binding regions. They proposed TATA-like elements to be the core promoter elements for ~98% of previously classified TATA-less promoters. The model that TATA-less promoters utilize TATA-like elements is intriguing and somewhat unifies all promoters in *S. cerevisiae*, however the TATA-like elements were not tested for their functionality in driving transcription in this study. The Hahn lab tested two TATA-less ribosomal protein promoters and showed that neither deletion of TATA-less elements, nor disruption of TBP DNA binding surfaces affected their transcription *in vitro* (Kamenova et al., 2014). The lack of effect of TATA-like element mutagenesis on transcription activity for one of promoters tested by Hahn group was consistent with an experiment by another group *in vivo* (Sugihara et al., 2011). We examined TATA-like elements relative to our high-resolution TSS data and tested several TATA-like elements for their functionality in Chapter IV.

Another possible core promoter element in TATA-less promoters is emerging from recent studies. The Cramer group reported a GA element (GAAAA) in 1/3 of TATA-less promoters at a stereotypical distance from TSS region, and the enrichment of GA element (GAE) anti-correlated with enrichment for consensus TATA elements (Seizl et al., 2011). They could also find the GAE in five related yeast species, although at a lower percentage of promoters than found in *S. cerevisiae*. In an experiment with a GAE-containing promoter, the GAE was able to drive transcription *in vitro*, confer TBP binding to an oligo *in vitro*, while promoter expression showed impairment when GAE itself or neighboring sequences were mutated. Other data suggesting that GAEs have some function for TATA-less promoters can be found in other publications as follows. Wu and Li reported strand specific enrichment of Poly-dA sequence led by a G (GAE) in the NFR regions of TATA-less promoters (Wu and Li, 2010); Zhang *et al* shows that GAA repeat sequences can drive expression of a promoter-less reporter gene *in vivo* in *S. cerevisiae,* and that Pol II initiation factors (TAFs and TFIIB) appear to control the ability of GAA repeats to be amplified during DNA replication (Zhang et al., 2012); This observation is consistent with them being bound by TFIID and TFIIB *in vivo*. Zentner and Henikoff identified a long GAE motif (a 6-7 mer GAE with 3-4 mer short G AE upstream and downstream) as being enriched in regions showing altered TBP occupancy in a *mot1* mutant (Zentner and Henikoff, 2013). Mot1 is a factor that can redistribute TBP from consensus TATA boxes to other locations in the genome. When Mot1 function was disrupted, TBP occupancy was reduced at other locations and increased at TATA boxes (Zentner and Henikoff, 2013). This suggested that GAE

regions were bound by TBP, consistent with function as alternative core promoter element for promoters in absence of TATA. We tested whether GAE could be a functional core promoter element at a few candidate TATA-less promoters in Chapter IV.

Existing model of TSS selection at different promoter classes in *S. cerevisiae*

In the study that proposed TATA-like elements to be the core promoter elements for TATA-less promoters, the authors proposed a model for TSS selection at different promoter classes (Rhee and Pugh, 2012). When TSSs in TATA-less/Taf1-enriched promoters were aligned relative to +1 nucleosome positions, it was shown that TSSs in this class were positioned at the border between the NFR/NDR and +1 nucleosome. Taf1 localization was more uniform in relation to TSSs and the +1 nucleosomes in the TATA-less/Taf1-enriched promoters. Conversely, TSSs in TATA-containing/Taf1-depleted promoters did not show apparent stereotypical positioning relative to the +1 nucleosome positions, even mapping deep inside the position of the annotated +1 nucleosomes. This led the authors to propose that in the TATA-less/Taf1-enriched promoter class, TAFs assemble cooperatively with the +1 nucleosome during formation of PIC, thus limiting Pol II scanning within the range of sequence upstream of the +1 nucleosome; while the TATA-containing/Taf1-depleted class promoters have PICs assemble competitively with the +1 nucleosome, leading to no constraint from +1 nucleosome in Pol II scanning (because the nucleosome would not be present) and allowing higher gene expression. The authors asked if annotated TSSs for TATA-containing promoters were on average further from the annotated TATA element than were TSS from their annotated TATA-

like element positions. They reported that aggregation of annotated TSS positions (David et al., 2006) relative to proposed TATA elements were indeed farther away compared to TSS positions from proposed TATA-like elements. We tested this model that essentially proposed that different promoter classes utilize different mechanisms for TSS selection/utilization by varied participation of +1 nucleosome with our own data in Chapter IV.

General Transcription Factors that have been shown to alter start site selection in *S. cerevisiae*

GTF mutants were the first discovered to alter TSS utilization at model genes in *S. cerevisiae* and therefore it was suggested that GTFs have critical roles in TSS selection. Much remains to be understood about the functions of GTFs and how they integrate with Pol II activity during initiation. Examination of mutant effects on TSS selection allows insight into the initiation process *in vivo* as TSSs are easily mapped and quantified. I first describe biochemical insights into two GTFs that have been suggested in many studies to play roles in TSS selection, followed by a description of alterations in TSS selection conferred by GTF mutants.

GTF TFIIB (encoded by *SUA7* in *S. cerevisiae*) has long been known as a critical coordinator of Pol II initiation and has been proposed to participate in TSS selection (Li et al., 1994). TFIIB bridges the TATA binding protein (TBP)- promoter DNA complex and Pol II, and likely stabilizes open complex formation through interaction with single-stranded DNA sequences in the PIC (Barberis et al., 1993; Buratowski and Zhou, 1993;

Nikolov et al., 1995). At the stage before the first phosphodiester bond of RNA is synthesized, TFIIB may stabilize template dsDNA and ssDNA in the open complex and play roles in TSS selection by Pol II. TFIIB can be segmented into four functional domains: B-ribbon, B-reader, B-linker and B-core cyclin folds. The B-ribbon interacts with the Pol II "dock" domain to recruit Pol II to promoter. The B-reader helix enters Pol II active site cleft near the RNA exit channel and is positioned adjacent to the Pol II "rudder" domain. TFIIB interaction with Pol II has been proposed to promote a conformational change of Pol II protrusion domain, affecting second Mg^{2+} atom binding to the Pol II active site (Kostrewa et al., 2009; Sainsbury et al., 2013). This second Mg^{2+} ion is one of two that are critical for Pol II catalytic activity (Steitz and Steitz, 1993; Svetlov and Nudler, 2013). Upon open complex conformation, the B-reader seems to be in contact with template DNA 7-8nt upstream from catalytic site, where it is proposed to stabilize the nascent DNA:RNA hybrid that forms once transcription initiates. When the nascent RNA grows to longer than ~6 bases, it should encounter the B-linker helix. Further chain elongation to 13 nucleotides of RNA will engender RNA clashes with the B-reader and B-ribbon, which is possibly a step that promotes the dissociation of TFIIB from Pol II (Pal et al., 2005).

The GTF TFIIF (encoded by conserved, essential genes *TFG1, TFG2,* and non-conserved and inessential gene *TFG3* in *S. cerevisiae*) enters the PIC after TBP and TFIIB bind to template DNA and may be recruited along with Pol II, guiding and stabilizing Pol II binding during assembly of the PIC. Furthermore, TFIIF appears to functionally interact with TFIIB (Chen and Hampsey, 2004; Chen and Hahn, 2004;

Eichner et al., 2010; Hampsey, 1998a; Henry et al., 1994; Luse, 2012; Sun and Hampsey, 1995). TFIIF has been proposed to interact with the Pol II protrusion domain, suggesting it may directly regulate Pol II activity, and with DNA both upstream and downstream of TFIIB B core domain (Miller and Hahn, 2006). Indeed TFIIF has been shown to alter Pol II catalytic activity in previous studies (Cabart et al., 2014; Fishburn and Hahn, 2012; Khaperskyy et al., 2008). A recent cryo-EM structure analysis suggests TFIIF opens the Pol II clamp (He et al., 2013). TFIIF and TFIIS have been shown to promote Pol II elongation by modifying Pol II pausing dynamics in single molecule studies (Grosso et al., 2012; Schweikhard et al., 2014), and shown to facilitate transcription through a nucleosome (Grosso et al., 2012).

Mutations in GTFs have previously been shown to alter TSS selection at a few model genes. Mutations in *SUA7* B-reader coding residues generally have been shown to alter TSS distribution towards downstream positions (downstream shift) (Chen and Hampsey, 2004; Faitar et al., 2001; Hull et al., 1995; Pinto et al., 1992; Pinto et al., 1994; Sun and Hampsey, 1995; Wu et al., 1999) although there are also upstream shift mutants reported (Zhang et al., 2002). When we reconstructed a subset of these upstream shifting *sua7* alleles, they conferred downstream shifts in our experiments consistent with all other reported *sua7* mutants (Chapter III, (Jin and Kaplan, 2014)). Conversely from the TFIIB alleles, a number of TFIIF *tfg1* and *tfg2* alleles have been shown to shift distribution of TSS towards upstream positions (Eichner et al., 2010; Freire-Picos et al., 2005; Ghazy et al., 2004; Hahn and Young, 2011; Khaperskyy et al., 2008; Majovski et al., 2005). Combinations of TFIIB and TFIIF alleles can confer mutual suppression of

their respective TSS defects along with TFIIF alleles' suppression of LOF TFIIB alleles' temperature sensitive phenotypes (Freire-Picos et al., 2005; Ghazy et al., 2004; Sun and Hampsey, 1995).

More recent studies have revealed alleles in another GTF TFIIH can also alter TSS utilization. Alleles of *SSL2*, which encodes an ATPase/helicase enzymatic subunit of GTF TFIIH (Grunberg et al., 2012), shifted the distribution of TSSs slightly upstream (Goel et al., 2012). TFIIH (comprising two ATPase/helicase subunits, the most important for initiation being Ssl2, a kinase subunit Kin28, and other 9 proteins) has roles in promoter melting, Pol II escape from the PIC and also in nucleotide excision repair (NER) (Dvir, 2002; Feaver et al., 1993; Jeronimo and Robert, 2014; Liu et al., 2004; Svejstrup et al., 1995; Wong et al., 2014). In my unpublished primer extension experiments of an *ssl2* allele tested previously by Hampsey and coworkers but without report of any TSS distribution change (Goel et al., 2012), ,I found that it conferred downstream shift in TSS distribution at *ADH1* using careful quantification of TSS usages. Another *ssl2* allele was shown to partially suppress downstream TSS shifts of *sua7-1* as well as *sua7-1* cold sensitivity (Goel et al., 2012). Ssl2 is a particularly interesting factor that affects TSS selection/utilization, being the only other protein subunit outside Pol II, which when mutated, shifts TSSs in both directions in allele-specific fashion. Our data presented in Chapter IV supports a hypothesis that Ssl2 is the driving force for translocating Pol II during TSS scanning from core promoter elements towards positions of TSSs.

Pol II alleles that alter TSS selection *in vivo*

Previously, alleles of some Pol II subunits have been shown to affect TSS usage distribution on their own and those of certain GTF alleles when Pol II and GTF alleles were combined. For a brief summary of factors affecting TSS selection, see **Table 1-2**. Mutations in Pol II subunit-encoding genes *RPO21/RPB1*, *RPB2, RPB7* and *RPB9* have been previously shown to alter TSS utilization *in vivo*; however the mechanisms of alterations have been unclear (Braberg et al., 2013; Chen and Hampsey, 2004; Chen et al., 2007; Freire-Picos et al., 2005; Hull et al., 1995; Kaplan et al., 2012; Sun and Hampsey, 1996; Sun et al., 1996). Combining *tfg1* and *rpo21/rpb1* alleles resulted both in suppressed temperature sensitivity and TSS defects of an *rpo21/rpb1* allele (Freire-Picos et al., 2005). Combining a *tfg1* allele and *rpb9*Δ resulted in exacerbated TSS defects and temperature sensitivity (Ghazy et al., 2004). Alleles of *rpb2* and *rpb9* were shown to suppress the downstream shift effect and cold temperature sensitivity of certain *sua7* alleles (Chen and Hampsey, 2004; Sun and Hampsey, 1996; Sun et al., 1996).

Table 1-2 Factors that are shown to alter TSS selection

Protein complex	Genes encoding protein subunit	Direction of shift
Pol II	*RPB1/RPO21, RPB2,RPB7,RPB9*	Upstream, downstream
TFIIB	*SUA7*	Downstream
TFIIF	*TFG1,TFG2*	Upstream
TFIIH	*SSL2*	Upstream, downstream

Our characterization of the relationship between Pol II catalytic activity mutants and TSS defects suggested an activity-based framework for interpretation of Pol II mutant TSS defects (Braberg et al., 2013; Kaplan et al., 2012). Through genetic experiments and examination of TSS selection *in vivo*, we have attempted to understand further the relationship between Pol II activity defects, GTF function, and transcription initiation in *S. cerevisiae* (Chapter II, III). Our preliminary data indicated TSS selection is altered *in vivo* in Pol II mutants with altered catalytic activity. In primer extension experiments that detect 5′ ends of transcript for the widely-used *ADH1* model for initiation, slow Pol II TL mutants showed downstream shifts of start sites similarly to most *sua7* alleles. Conversely, fast Pol II mutants showed upstream shifts of *ADH1* start sites similarly to *tfg2* alleles and *rpb9*Δ. A Pol II double mutant combining slow and fast mutations, which shows suppression of both single mutant's activity defects (thus has activity similar to WT), also showed suppression of the TSS selection defects seen in the

individual slow and fast mutants. The severity of TSS defects *in vivo* in both fast and slow mutants as well as the double mutant correlated well with the extent of their deviation from WT elongation rate *in vitro*. With the observation that these TSS defects closely correlated with altered catalytic activity, we proposed that altered catalytic activity affects start site selection directly (Kaplan et al., 2012). These results raised questions regarding why catalytic activity will influence distribution of TSS, as selection of the TSSs might be considered an event preceding occupancy of the Pol II active site and rate of phosphodiester bond formation. A mechanistic explanation for the connection of Pol II activity and TSS selection will be critical for understanding Pol II initiation (Chapter IV).

Large scale study of TSS selection in *S. cerevisiae*

The first semi-high throughput sequencing of TSSs used 5′ SAGE (Serial Analysis of Gene Expression) sequencing to map TSSs for ~30% genes in *S. cerevisiae* (Zhang and Dietrich, 2005). 5′ SAGE and CAGE (Cap Analysis Gene Expression) sequencing are widely used sequencing methods in which mRNAs are selected by Poly-A tails at their 3′ ends and/or their 5′ cap structure. They have one flaw that can affect the precise determination of the 5′ RNA end. In these methods, mRNAs are reverse transcribed to generate cDNA before being ligated with an adaptor sequence unlike other more recently developed methods; because Reverse Transcriptase potentially adds extra Cs complementary to the m^7G cap structure, the resulting sequence of mRNA can contain extra Gs (Plessy et al., 2010). Since an initiating nucleotide can be a G, it is hard

to determine if the initiating nucleotide is G or is the extra G added on. In Zhang and Dietrich's analyses, all Gs at 5′ of reads were removed because 5′ adaptor sequence had 3′ GGG to remove overrepresentative 5′ G due to adaptor sequences. In the case where G was upstream nucleotide of the stripped sequencing read, a G was added at 5′, *i.e.* G at 5′ in this study was sub-optimally handled. With this in mind, the initial SAGE data also did not offer very deep read coverage over the whole genome. However, these data provided valuable insights into TSS utilization. First, TSSs at 80% of TATA-containing promoters agreed with previous report that TSSs in *S. cerevisiae* are ~40-120 nt downstream from the TATA element, confirming *S. cerevisiae* TSSs being relatively distal to the TATA box compared to other eukaryotes for TATA-containing promoters. Second, ~5000 TSSs showed the usage of purine (R, A/G) as the initiating nucleotide (+1), while pyrimidine (Y, C/T) was the upstream neighbor nucleotide (−1) with an additional dominant preference for A at the −8 position relative to the initiation position. It is worth noting that this data set mainly captures the TSSs of the promoters with the highest expression, thus the TSS motifs discovered in this study represent the promoters that have high output (expression) (Zhang and Dietrich, 2005).

More recently, TSS mapping with deeper read coverage and more accuracy have been published by other groups (Kang et al., 2015; Malabat et al., 2015; Park et al., 2014; Pelechano et al., 2013; Rojas-Duran and Gilbert, 2012). Naturally each method has drawbacks and biases of its own. Errors and biases that can occur in large scale genome wide studies include sequencing errors, GC-biases in PCR amplification, ligation bias in library construction, and read mapping ambiguity due to low confidence reads. Many

genome wide studies mentioned above employ simultaneous mapping of 5′ and 3′-RNA by linking ends of poly-A selected RNA, thus bias against transcripts that are not poly-A enriched or are difficult to fully reverse transcribe. We wished to examine influence of Pol II activity in TSS selection and other initiation properties *in vivo*, therefore we utilized SOLiD 5′-RNA sequencing which has been shown as a method with a high accuracy to map TSS and investigated Pol II activity-based assessment of TSS selection/utilization and control by promoter architecture in global scale (Chapter IV).

Models for TSS selection in eukaryotes

TSS selection in *S. cerevisiae* has been considered to be mechanistically different from higher eukaryotes due to the relatively large distance from TATA elements to TSSs in *S. cerevisiae* and the perception that TSS distributions were wider than biochemically studied human promoters (reviewed in (Kaplan, 2013). A powerful model that is consistent with most, if not all, existing data in *S. cerevisiae* is that Pol II scans for usable TSSs from the core promoter element (**Figure 1-3** A). In the following section, I briefly summarize the key evidence that supports the scanning mechanism in *S. cerevisiae* and give a brief review of supportive literature. Giardina and Lis used permanganate footprinting to demonstrate that a wide region of in *S. cerevisiae* promoters upstream of TSSs was melted, indicating that the melting region started at a point similar to human promoters (Giardina and Lis, 1993). Additionally, Kuehner and Brow demonstrated that when multiple preferred start sites were placed within optimal distance from core promoter, the upstream starts were preferably used (“first come, first

served"), yet mutation at upstream sites led to usage of downstream usable start sites (Kuehner and Brow, 2006). Third, genome wide GTFs mapping by Rhee and Pugh showed GTFs localization upstream of TSSs indicating that GTFs would need to shift to localize to actual sites of initiation (Rhee and Pugh, 2012). Fourth, mutants in a number of factors that alter start site usage do so in a polar, directional fashion as mentioned above, with different Pol II and *ssl2* (helicase subunit of TFIIH) alleles able to cause shifts in either direction in an allele-specific fashion.

How Pol II scans for TSS is an open question. It was shown that Pol II can initiate transcription from downstream start sites in the presence of a chain terminating nucleotide 3′-O-methyl GTP, which should block chain elongation between the promoter and downstream start sites at all G positions, suggesting that continuous transcription is not required for scanning (Khaperskyy et al., 2008; Lue et al., 1989). How the output of scanning is controlled and what are the main determinants of scanning properties for yeast promoters are not known (discussed in Chapter IV). As more recent studies revealed more similarities of yeast promoters with higher eukaryotes (*i.e.* majority of promoters specifying multiple TSS) (Chen et al., 2013; Consortium et al., 2014; Haberle et al., 2014; Hoskins et al., 2011; Li et al., 2015), yeast has become the perfect model system to shed light on Pol II mechanisms during initiation and the regulation.

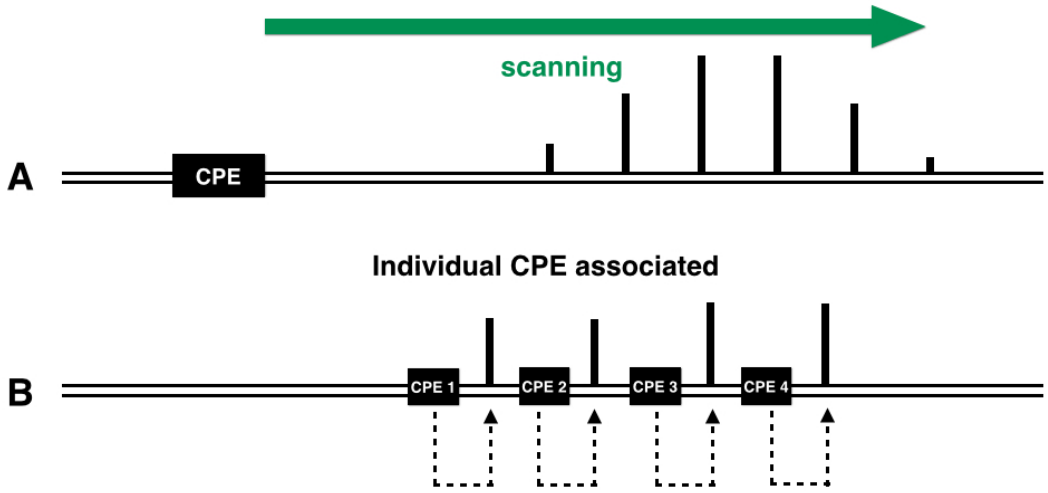


Figure 1-3. Models for dispersed TSS selection.
A. Pol II is recruited to core promoter element (CPE) can scan directionally for TSSs.
B. Each TSS has individually associated CPE upstream driving each site of dispersed TSSs.

Another model for multiple dispersed TSSs is that they arise from the presence of multiple core promoter element clusters. Some promoters in other eukaryotes have been suggested to utilize this mechanism, for example, promoters for maternal transcripts in zebrafish have a WW (W=A/T) box ~30nt upstream associated with each TSS (Haberle et al., 2014) (**Figure 1-3** B). Relatively little is known about individually associated core promoter elements in eukaryotes and what fraction of metazoan promoters are in this class. A high resolution of GTF mapping in other eukaryotes may aid search of such core promoter elements.

I have taken a bioinformatics approach to compare usage of initiator $Y_{-1}R_{+1}$ motifs in promoters of other eukaryotes (Chapter IV). These analyses are able to differentiate promoters where TSSs distributions match those in yeast and thus have properties consistent with scanning or those that share characteristics with zebrafish maternal promoters, and may have TSSs individually associated with core promoter element clusters (**Figure 1-3**). Our preliminary analyses support a hypothesis that at least a subset of higher eukaryotic promoters are scanned for TSS selection (Chapter IV).

Literature that supports Pol II scanning mechanism for transcription start site selection in *S. cerevisiae*

The model that Pol II scans for TSS was first proposed by Giardina and Lis (Giardina and Lis, 1993). Using permanganate (KMnO4) footprinting that detects single stranded thymidine, they detected promoter melting at model genes *GAL1* and *GAL10*. Although two genes have different distances between TATA element and TSS that are

within a standard range of *S. cerevisiae* promoters, they both showed melting that started ~20nt downstream of TATA elements thus different widths of melting region depending on their distances to TATA elements. This places the melting region of these model genes at positions similar to those of other eukaryotes (Buratowski et al., 1991; Giardina et al., 1992; Wang et al., 1992). The authors proposed that Pol II is recruited to the core promoter element, melts the promoter region and scans downstream for the TSSs (Giardina and Lis, 1993).

The idea of Pol II scanning for TSS can also be seen in (Chen and Struhl, 1985) in which work authors used mutants with altered promoter distance between TATA element and TSS at a model gene as well as promoter swap mutants and tested TSS selection in these mutants. They concluded that local TSS motifs rather than a certain distance from core promoter element were the determinant for use of TSSs. They utilized a chimeric construct with core promoter section of a gene and TSS region of another gene. Although transcript levels were determined by the core promoter element region sequences, TSS utilization was consistent with the gene to which the TSS segment belonged. Used TSSs in chimeric promoters were at different distances from the original core promoters for which the TSSs belonged, but were dependent on the core promoter driving the chimeric promoters. Although it seemed the authors understated the importance of distance between core promoter and TSS, this work hinted at the idea of TSS motif preference determining TSS positioning rather than TSS positioning arising as an outcome of a core promoter element-determined architecture.

An enlightening discussion can also be found in (Furter-Graves and Hall, 1990). The authors introduced an *S. pombe* gene and its mutants with varied distances between TATA element and TSS into *S. cerevisiae* to investigate determinants for TSS sites in *S. cerevisiae*. The authors found that while core promoter element required a minimal distance between TSS and CPE, the CPE was "not required to discriminate suitable from unsuitable sites". When strong TSS motifs were placed within the optimal distances from TATA, yet upstream of the most used TSSs in WT, the upstream TSSs were used preferably, and led to reduced usage of downstream TSSs. The authors hypothesized that "*search* for suitable [initiation] sites" within an initiation window proceeds in a 5′ → 3′ direction (Furter-Graves and Hall, 1990).

A more comprehensive and quantitative support for Pol II scanning initiation in *S. cerevisiae* initiation was presented by (Kuehner and Brow, 2006). The authors used a model gene, *SNR14*, which has a strongly preferred TSS, and variants including mutants with altered TATA-TSS distances, and those that have tandem repeats of a short TSS region (including the strong TSS). Preferred usage of the upstream TSS repeat in a synthetic promoter with repeated TSS sequences was observed, and interestingly, when the upstream strong TSS was mutated to a weak TSS, the downstream TSS became the dominant TSS. These results supported 5′ → 3′ directional and processive scanning for TSS as well as each TSS having an intrinsic efficiency determined by local sequence context (Kuehner and Brow, 2006).

All the mutants that have been shown to alter TSS distribution at model genes (described in previous section) do so in a directional manner, supporting the directional

scanning mechanism. So far the supportive studies of scanning have focused on TATA-containing promoters in *S. cerevisiae*, however the majority of *S. cerevisiae* promoters are TATA-less and have not been tested for scanning. I present our recent findings in Chapter IV that further supports the scanning mechanism, both for TATA-containing and TATA-less promoter classes and discuss the possibility that a set of other eukaryotic promoters may be scanned for TSS in a manner similar to *S. cerevisiae.* These studies challenge the previous suggestion that mechanism of TSS selection in *S. cerevisiae* is fundamentally distinct from other eukaryotes (Corden et al., 1980; Hampsey, 2006; Russell, 1983).

CHAPTER II

RELATIONSHIP BETWEEN ALTERED POL II ACTIVITY AND TRANSCRIPTION START SITE UTILIZATION*

Overview

Chapter II comprises reprints of two publications with permissions to which I contributed, being second author on both. I describe my contributions in disclaimers below. Summaries are reprints of the abstracts of the publications; the rest of the sections are as published.

Disclaimer for first sub-chapter

In a preliminary primer extension experiment by Dr. Kaplan that detected RNA 5′ ends of *ADH1*, it was observed that Pol II TL mutants had polar effects on start site selection at *ADH1*—a low activity mutant (LOF) had downstream shifts on TSS selection, while a high activity mutant (GOF) had upstream shifts on TSS selection. I performed primer extension experiments on additional model genes using an expanded mutant set to assess additional TL mutants with known *in vitro* elongation activity effects and genetically related Pol II mutants outside the TL. The effects observed were

consistent based on predictions for mutant class (GOF/LOF) based on *in vitro* activity or *in vivo* genetic phenotypes. The magnitudes of TSS shifts in model gene *ADH1* were especially interesting because they correlated closely with the magnitude of elongation rate defects (whether faster or slower than WT). TSS detection at *GAL1* was particularly interesting- I discovered that fast Pol II mutants utilize previously unseen upstream TSSs within the region where Giardina and Lis detected promoter melting, close to the upstream edge of melting region, supporting the scanning model. I also conducted quantitative western blot experiments that showed that abundance of Rpb1 in mutants was not strongly altered from WT. Figure contribution in this publication: Figure 6, Figure S6, Figure S7 in (Kaplan et al., 2012)).

Disclaimer for second sub-chapter

I conducted growth phenotyping experiments for Pol II mutants used in high throughput pE-MAP (point mutant epistatic miniarray profile) and verified that pE-MAP was able to dissect function of different domains of complex proteins at single residue resolution. I determined relationships between start site defects and conditional growth phenotypes of Pol II mutants by correlating the degree of start site defects observed at the model gene *ADH1* and Spt$^-$ (Suppressor of Ty) and mycophenolic acid (MPA) sensitive phenotypes. I found that the MPA phenotype correlates more than the Spt$^-$ phenotype with TSS defects; specifically upstream shifts at *ADH1* correlate with MPA sensitivity. It indicates that *IMD2* (the gene responsible for determination of MPA sensitivity) and *ADH1* may be sensitive to similar types of TSS defects, while the Spt$^-$

phenotype, which relies on altered transcription of *lys2-128∂*, if affected by TSS defects, is affected in different way than *IMD2* and *ADH1*. I present evidence supporting that pE-MAP analysis is able to cluster Pol II mutants that show different phenotypes by testing their conditional growth phenotypes. The pE-MAP can be a high throughput measurement for probing phenotypes that originate from perturbing different aspects of essential processes, for instance transcription. This also supports our model that the phenotypes that correlate with start site defects may actually be caused by a start site defects. In addition, I conducted the growth phenotyping studies to verify E-MAP's ability to detect genetic interactions of Pol II mutants and a large number of tested and putative transcription factors mutants. I studied *SUB1* (SUppressor of TFIIB), a gene that shows wide-ranging allele-specific genetic interaction with Pol II mutants in the pE-MAP, by generating *sub1*Δ and determining its modulation of TSS effects and conditional growth phenotypes of Pol II mutants. I showed that our conditional growth phenotyping assay reflects a wide range of genetic interaction, consistent with our pE-MAP results. I discovered *sub1*Δ alters TSS utilization at *ADH1*, and also affects TSS phenotypes of Pol II mutants, which mirror genetic interactions between *sub1*Δ and both classes of Pol II mutants. Figure contribution to this publication: Figure 1, Figure 4, Figure 7, Figure S1, Figure S4, Figure S6, Figure S7 in (Braberg et al., 2013).

Dissection of Pol II Trigger Loop function and Pol II activity–dependent control of start site selection *in vivo*

Summary

Structural and biochemical studies have revealed the importance of a conserved, mobile domain of RNA Polymerase II (Pol II), the Trigger Loop (TL), in substrate selection and catalysis. The relative contributions of different residues within the TL to Pol II function and how Pol II activity defects correlate with gene expression alteration *in vivo* are unknown. Using *Saccharomyces cerevisiae* Pol II as a model, we uncover complex genetic relationships between mutated TL residues by combinatorial analysis of multiply substituted TL variants. We show that *in vitro* biochemical activity is highly predictive of *in vivo* transcription phenotypes, suggesting direct relationships between phenotypes and Pol II activity. Interestingly, while multiple TL residues function together to promote proper transcription, individual residues can be separated into distinct functional classes likely relevant to the TL mechanism. *In vivo*, Pol II activity defects disrupt regulation of the GTP-sensitive *IMD2* gene, explaining sensitivities to GTP-production inhibitors, but contrasting with commonly cited models for this sensitivity in the literature. Our data provide support for an existing model whereby Pol II transcriptional activity provides a proxy for direct sensing of NTP levels *in vivo* leading to *IMD2* activation. Finally, we connect Pol II activity to transcription start site selection *in vivo*, implicating the Pol II active site and transcription itself as a driver for start site scanning, contravening current models for this process.

Introduction

Cellular DNA-dependent RNA polymerases likely balance fidelity in substrate selection with synthesis speed to achieve appropriate transcriptome content and regulation *in vivo*. In multisubunit RNA polymerases (msRNAP) from archaea, bacteria and eukaryotes, a highly conserved subdomain known as the trigger loop (TL) is critical for rapid catalysis and selection of correct substrates (Kaplan, 2010; Kaplan et al., 2008; Kireeva et al., 2008; Temiakov et al., 2005; Toulokhonov et al., 2007; Vassylyev et al., 2007; Wang et al., 2006; Yuzenkova et al., 2010; Zhang et al., 2010). The TL is present in the largest subunit of eukaryotic Pol II, generally referred to as Rpb1 (Rpo21 in *Saccharomyces cerevisiae*), and the analogous b′ subunit of bacterial RNAP, and A″ subunit of archaeal RNAP.

Similarly to mobile domains of other classes of nucleic acid polymerases, the TL undergoes conformational changes in conjunction with the presence of an NTP substrate complementary to the DNA template (matched) in the msRNAP active site (Vassylyev et al., 2007; Wang et al., 2006). These conformational changes are proposed to link TL-substrate interactions to preferential catalysis of correctly matched substrates over mismatched substrates. The TL can be observed in distinct conformations depending on the presence of matched NTP substrate, natural product polymerase inhibitors, and msRNAP-interacting proteins, underscoring its flexibility (Brueckner and Cramer, 2008; Cheung et al., 2011; Kaplan et al., 2008; Kettenberger et al., 2003; Spahr et al., 2009; Sydow et al., 2009; Tagami et al., 2010; Vassylyev et al., 2007; Wang et al., 2006). In addition to effects on phosphodiester bond catalysis, the TL has been implicated in

polymerase pausing, intrinsic cleavage of RNA and translocation (Bar-Nahum et al., 2005; Kireeva et al., 2008; Toulokhonov et al., 2007; Yuzenkova and Zenkin, 2010; Zhang et al., 2010).

The TL comprises two mostly alpha-helical regions connected by a short loop (Figure 1A). Deletion or structural compromise of the TL in either *E. coli* (*Eco*) or *T. thermophilus* (*Tth*) strongly reduces catalytic activity, but to different extents depending on whether the substrate is matched to the template or not (Temiakov et al., 2005; Vassylyev et al., 2007; Yuzenkova et al., 2010; Zhang et al., 2010). Two specific regions of the TL appear important for control of TL function. First, conserved residues in the nucleotide interacting region (NIR) recognize specific features of matched NTP substrates and work in conjunction with non-TL residues, Rpb1 N479 and R446, positioned for interaction with hydroxyl moieties on the ribose of matched NTPs (Vassylyev et al., 2007; Wang et al., 2006)(**Figure 2-1** A). Residues in the NIR showing NTP interactions in *S. cerevisiae* and *Tth* substrate-bound structures are (using *S. cerevisiae* Rpb1/Rpo21$^{Eco\beta'/Tth\beta'}$ numbering) Gln1078$^{Gln930/Gln1236}$, Leu1081$^{Met932/Met1238}$, Asn1082$^{Arg933/Arg1239}$ and His1085$^{His936/His1242}$. Second, in all kingdoms of life, substitutions in or near the helix distal to the NIR alter elongation rate, in some cases increasing elongation rate relative to WT ("superactivity"), (Bar-Nahum et al., 2005; Kaplan et al., 2008; Malagon et al., 2006; Tan et al., 2008). These substitutions may alter dynamics of TL movement between the substrate-interacting conformation and other conformations because they are adjacent to the hinge region in the C-terminal TL helix

(another hinge is apparent in the N-terminal TL helix)(**Figure 2-1** B) (Kaplan et al., 2008; Kireeva et al., 2008).

NIR residues of the *Eco* and *Tth* RNAPs have different degrees of contribution to catalytic activity, with individual *Eco* residue substitutions having smaller effects on activity than homologous substitutions in *Tth*, underscoring the diversity of TL residue functions in msRNAPs (Yuzenkova et al., 2010; Zhang et al., 2010). A conserved histidine within the TL NIR has been proposed to have an important or essential catalytic role (Wang et al., 2006). Many classes of structurally distinct nucleic acid polymerases utilize a basic residue located at a position in the active site analogous to that of the TL histidine and these residues function as a general acid in the catalytic cycle of the polymerase (Castro et al., 2009). While clearly important for RNAP catalysis, the TL histidine appeared not to be functioning as a general acid based on pH-dependence curves for bacterial RNAPs (Yuzenkova et al., 2010; Zhang et al., 2010). Experiments in the different bacterial systems show that a TL methionine residue (equivalent to *S. cerevisiae* Leu1081) packs against a base-paired NTP in the active site and has a greater contribution to activity than $Arg^{Eco933/Tth1239}$ or $His^{Eco936/Tth1242}$, which interact with the triphosphate moiety of matched substrates (Yuzenkova et al., 2010; Zhang et al., 2010).

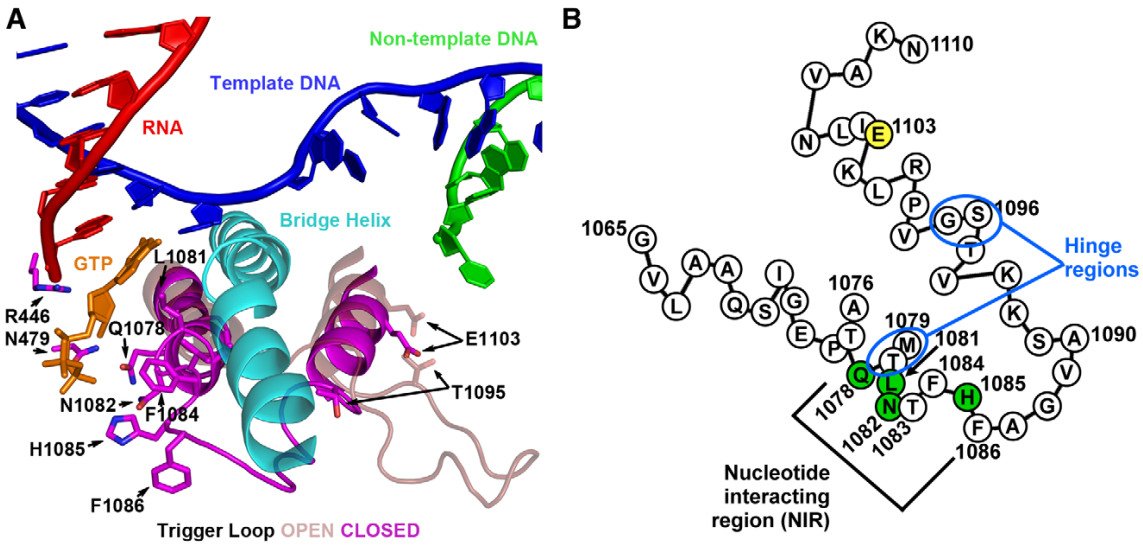


Figure 2-1. *S. cerevisiae* Rpb1 trigger loop conformations and sequence.
A. Cartoon representation of "closed" Pol II TL in relation to nucleic acids, Rpb1 bridge helix and a matched GTP substrate from structure PDB 2E2H (Wang et al., 2006) overlaid with TL constrained in open conformation by TFIIS (not shown) from structure PDB 1Y1V (Kettenberger et al., 2004). Amino acids (all derived from Rpb1) adjacent to the matched GTP substrate are indicated by numbers and single-letter amino acid codes. This figure was created with Pymol (Schrodinger, 2010). **B.** Schematic of TL showing amino acid sequence in single-letter code, with positions of interest numbered, residues with direct contact to GTP substrate in structure 2E2H shown in green, and position 1103 shown in yellow. Two hinge regions, about which the TL appears to change conformation from the open to closed positions are indicated.

Genetic identification of mutants with substitutions in the TL region, first in *S. cerevisiae* Pol II, then in *Eco* RNAP, demonstrated that alteration of the TL could alter transcription *in vivo* (Archambault et al., 1998; Bar-Nahum et al., 2005; Hekmatpanah and Young, 1991; Malagon et al., 2006; Weilbaecher et al., 1994). How changes in TL function or msRNAP catalysis alter transcription *in vivo* are not well understood and to what extent polymerase activity defects may be tolerated *in vivo* is not clear. In *S. cerevisiae*, drugs that target nucleotide synthesis pathways such as mycophenolic acid (MPA, targets GTP synthesis) (Sweeney, 1977) and 6-azauracil (6-AU, targets UTP and GTP synthesis) (Exinger and Lacroute, 1992; Handschumacher and Pasternak, 1958) have been shown to cause alterations in gene expression *in vivo* (Grigull et al., 2004; Howe et al., 2003; Kulish and Struhl, 2001; Mason and Struhl, 2005). A large number of Pol II transcription-related mutant strains show altered sensitivities to these drugs, leading to the broadly utilized interpretation that these drugs are elongation inhibitors and that sensitivity to them suggests defective Pol II elongation (Archambault et al., 1992; Desmoucelles et al., 2002; Du et al., 2008; Ginsburg et al., 2009; Hartzog et al., 1998; Hemming et al., 2000; Keogh et al., 2003; Kim et al., 2004; Lindstrom and Hartzog, 2001; Malagon et al., 2006; Orphanides et al., 1999; Otero et al., 1999; Pascual-Garcia et al., 2008; Powell and Reines, 1996; Reines, 2003; Riles et al., 2004; Schneider et al., 2010; Schwabish and Struhl, 2007; Shaw and Reines, 2000; Shaw et al., 2001; Song and Ahn, 2010; Sun et al., 2010; Wittschieben et al., 1999; Wu et al., 1996; Xiao et al., 2005). Notwithstanding the large number of mutants sensitive to these drugs that have no known transcriptional role, many MPA-sensitive mutants alter transcription

of the gene *IMD2* (Desmoucelles et al., 2002; Riles et al., 2004; Shaw and Reines, 2000), which encodes an MPA-resistant form of IMPDH, the enzymatic activity targeted by the drug (Hyle et al., 2003; Jenks and Reines, 2005). This gene-specific transcription defect is not always considered when interpreting mutant phenotypes. Intriguingly, *IMD2* transcription involves a switch between upstream transcription start site and downstream productive start sites that differ in initiating NTPs (upstream: GTP, downstream: ATP) leading to the proposal that the initiation process for these different classes of transcript stems from GTP levels being sensed directly by Pol II (Jenks et al., 2008; Kuehner and Brow, 2008).

The eukaryotic Pol II system provides an excellent model for *in vivo* studies of how the TL functions in transcription. Because nuclear transcription in eukaryotes is segregated among three essential polymerases instead of one, as in bacteria and archaea, strong defects may be more tolerated in Pol II than bacterial or archaeal RNAPs *in vivo*. We utilized extensive site-directed mutagenesis of the Pol II TL coupled with genetic screening to identify viable substitutions in TL NIR residues and substitutions conferring conditional growth phenotypes. We then used biochemical characterization together with a number of *in vivo* genetic and molecular phenotypes to probe the contributions of critical TL residues to transcription *in vivo*. We determined the relationship between NIR residues and superactivating TL substitutions and found that superactivating substitutions were mostly mutually suppressive with loss-of-function substitutions within the NIR in genetic and biochemical assays. These results indicated that NIR residues were not bypassed by superactivating substitutions and were still required for

Pol II activity. Using a series of TL mutants that define a continuum of *in vitro* elongation rates, we demonstrated that a number of *in vivo* phenotypes correlated closely with Pol II activity *in vitro*. We provide support for models proposing that *IMD2* transcription is directly sensitive to Pol II activity, therefore explaining the MPA-sensitivity of superactive Pol II mutants, which otherwise might have been expected to be resistant to reduced GTP levels due to increased elongation activity. Finally, we determined that start site selection at a number of other genes was similarly sensitive to alteration in Pol II activity leading to a new model for transcription-dependent polarity of start site selection in *S. cerevisiae*.

Results

Genetic Analyses Of Rpb1 TL Mutants In Vivo

We have undertaken an extensive genetic dissection of the Pol II TL: specifically, we examined the contribution of TL residues to Pol II function, and how Pol II catalytic activity relates to transcription *in vitro* and *in vivo*. To examine a large number of Pol II mutants *in vivo*, we employed a yeast strain containing a deletion of the endogenous copy of *RPO21* (the gene encoding Rpb1, which we henceforth refer to as *RPB1*) with *RPB1* activity complemented by a low copy *CEN* plasmid containing *RPB1* genomic DNA or mutant variants, allowing expression from the native *RPB1* promoter. Site-directed mutagenesis was focused on TL NIR residues to identify viable substitutions and was combined with existing *rpb1* TL alleles identified in genetic screens. For any viable mutants, alteration in transcription *in vivo* was measured with two phenotypic

reporters for transcriptional defects, the *lys2-128∂* (Simchen et al., 1984) and *gal10Δ56* alleles (Greger and Proudfoot, 1998; Kaplan et al., 2005)(**Figure 2-2**). These reporters are modulated by a number of transcription elongation factors and Pol II mutants with known transcription defects (Bucheli and Buratowski, 2005; Bucheli et al., 2007; Kaplan et al., 2005; Kaplan et al., 2008; Malone et al., 1991; Prelich and Winston, 1993; Winston et al., 1984).

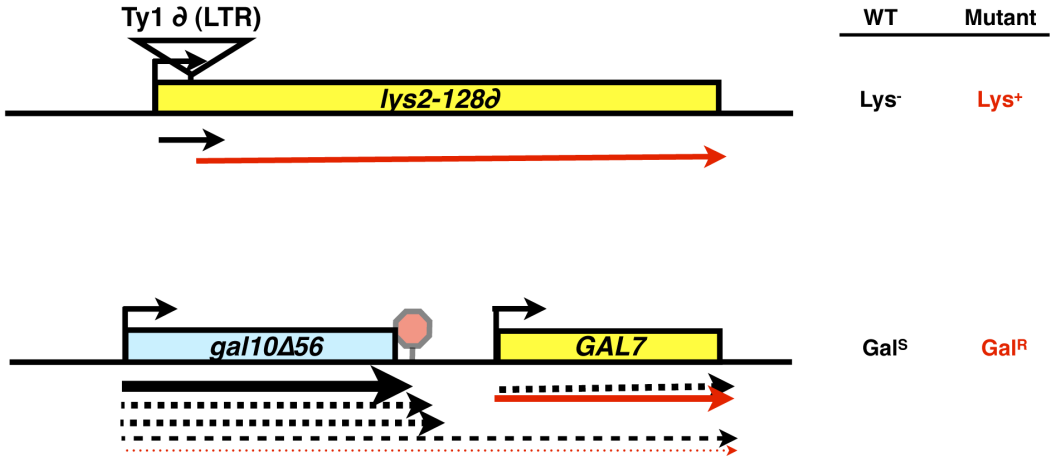


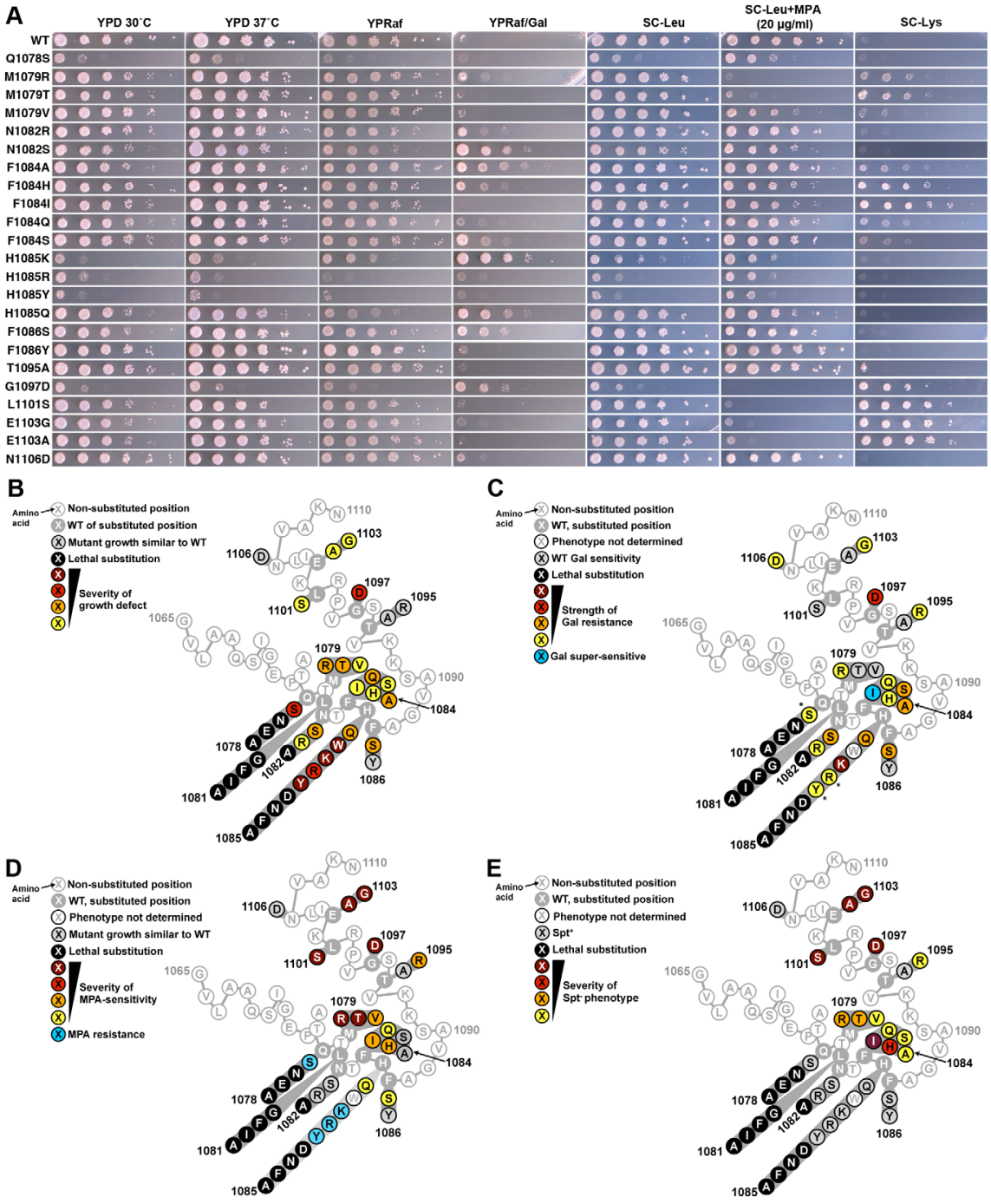
Figure 2-2. *In vivo* transcriptional phenotypes utilized in this study.
A. Cartoon representation of "closed" Pol II TL in relation to nucleic acids, Rpb1 bridge helix and a matched GTP substrate from structure PDB 2E2H (Wang et al., 2006) overlaid with TL constrained in open conformation by TFIIS (not shown) from structure PDB 1Y1V (Kettenberger et al., 2004). Amino acids (all derived from Rpb1) adjacent to the matched GTP substrate are indicated by numbers and single-letter amino acid codes. This figure was created with Pymol (Schrodinger, 2010). **B.** Schematic of TL showing amino acid sequence in single-letter code, with positions of interest numbered, residues with direct contact to GTP substrate in structure 2E2H shown in green, and position 1103 shown in yellow. Two hinge regions, about which the TL appears to change conformation from the open to closed positions are indicated.

In our current study, we found that the Spt$^-$ phenotype was concentrated in residue substitutions proximal to the two TL "hinges" (**Figure 2-3** A, E, see **Figure 2-1** B for positions of hinges). Growth defects were conferred by substitutions throughout the TL (**Figure 2-3** A, B). We found that MPA-sensitivity (MPA^s) and Spt$^-$ phenotypes generally co-occurred although the relative strength of the two phenotypes varied among mutants (**Figure 2-3** A, D, E). Finally, we found that the strongest suppressors of *gal10*Δ*56*, with the exception of G1097D, were in TL NIR residues and were mostly Spt$^+$ or MPA-resistant (MPA^r), indicating a distinction between these phenotypes (Figure **2-3** C). Taken together, these conditional plate phenotypes behave as sensitive readouts for likely complex sets of overlapping or distinct transcriptional defects *in vivo*.

In Vitro Transcription Elongation Rates Correlate With In Vivo Growth Defects And Plate Phenotypes

As TL substitutions causing the Spt$^-$ phenotype included those with known TL GOF phenotypes (F1084I, E1103G) (Kaplan et al., 2008), we asked if our plate phenotypes were predictive of biochemical phenotypes. Analysis of TL substitutions in an *in vitro* transcription assay using reconstituted Pol II elongation complexes indicated a good correlation between elongation rate defect *in vitro* and growth defect *in vivo* (**Figure 2-4**, **Figures 2-5** and **Figure 2-6**). We note that we did not detect strong expression differences for tested mutant Rpb1 proteins (**Figure 2-7**). We previously

Figure 2-3. Genetic analyses of Pol II TL single substitution mutants.
10-fold serial dilutions of saturated cultures of Pol II TL mutant strains plated on different media. YPD is rich medium with dextrose as a carbon source. YPRaf is rich medium with raffinose as a carbon source. YPRaf/Gal has both raffinose and galactose, allowing assay of *gal10*Δ*56*-dependent galactose toxicity phenotypes. SC-Leu is defined, complete medium lacking leucine. MPA was added to this medium (SC-Leu+MPA) to 20 μg/ml final concentration, showing that a number of Pol II mutants are sensitive to this drug. SC-Lys is defined, complete medium lacking lysine, and detects the Spt^- phenotype (Lys^+) for strains containing *lys2-128∂*. WT strains grow robustly on most media, but will not grow on SC-Lys when *lys2-128∂* is present and grow very poorly on YPRaf/Gal when *gal10*Δ*56* is present. Mutant-dependent transcriptional phenotypes allow modulation of these specific growth defects. **B.** Schematic of TL (as diagrammed in Figure 2-1 B) showing distribution of viable substitutions with their growth defects on YPD as well as lethal substitutions. All mutants were examined in the background of an Rpb1 T69 change, which is the allele present in the S288C reference genome as well as being normally found in our strains, representing a distinction for the subset of mutants described previously (Kaplan et al., 2008)(see Materials and Methods, Text S1 and note concerning viability of Q1078S). Scoring of phenotypic strength is based on visual inspection of (A). **C.** Distribution of *gal10*Δ*56* suppression (Gal resistance) or enhancement (Gal super-sensitivity) among TL substitutions. Scoring of phenotypic strength is based on visual inspection of (A). *Indicates strength of *gal10*Δ*56* suppression is likely underestimated due to confounding generic growth defects. **D.** Distribution of MPA phenotypes among TL substitutions. Scoring of phenotypic strength is based on visual inspection of (A). **E.** Distribution of *lys2-128∂* suppression (Spt^- phenotype scored as Lys^+) among TL substitutions. Scoring of phenotypic strength is based on visual inspection of (A).


A
YPD 30˚C
YPD 37˚C
YPRaf
YPRaf/Gal
SC-Leu
SC-Leu+MPA (20 µg/ml)
SC-Lys
WT
Q1078S
M1079R
M1079T
M1079V
N1082R
N1082S
F1084A
F1084H
F1084I
F1084Q
F1084S
H1085K
H1085R
H1085Y
H1085Q
F1086S
F1086Y
T1095A
G1097D
L1101S
E1103G
E1103A
N1106D
B
Amino acid
Non-substituted position
WT of substituted position
Mutant growth similar to WT
Lethal substitution
Severity of growth defect
C
Amino acid
Non-substituted position
WT, substituted position
Phenotype not determined
WT Gal sensitivity
Lethal substitution
Strength of Gal resistance
Gal super-sensitive
D
Amino acid
Non-substituted position
WT, substituted position
Phenotype not determined
Mutant growth similar to WT
Lethal substitution
Severity of MPA-sensitivity
MPA resistance
E
Amino acid
Non-substituted position
WT, substituted position
Phenotype not determined
Spt+
Lethal substitution
Severity of Spt- phenotype

showed that LOF substitution H1085Y and a putative GOF substitution (G1097D) caused severe growth defects *in vivo* (Kaplan et al., 2008). In this study we find that these two mutants also represent extremes of Pol II activity defects, but interestingly at different ends of the activity spectrum. Biochemical analysis of Spt^- and MPA^S G1097D and L1101S shows that they are GOF substitutions conferring increased maximal elongation rate *in vitro*. In fact, G1097D elongates too quickly for accurate measurement in our short template run off assay (**Figure 2-4, Figure 2-5** and **Figure 2-6**). We evaluated a number of other TL substitutions and found that WT elongation rate *in vitro* correlates with robust growth *in vivo,* and that the greater the deviation from WT for elongation rate *in vitro*, the greater the deviation from WT growth *in vivo* (**Figure 2-4**, right panel). We note that the non-TL LOF substitution, N479S, deviates slightly from this relationship as it has the most severe *in vitro* defect but has a less severe growth defect. This could be consistent with a factor(s) functioning *in vivo* that can compensate for the biochemical defect caused by the N479S substitution but not the H1085Y substitution; for example, a hypothetical factor that may work through the TL.

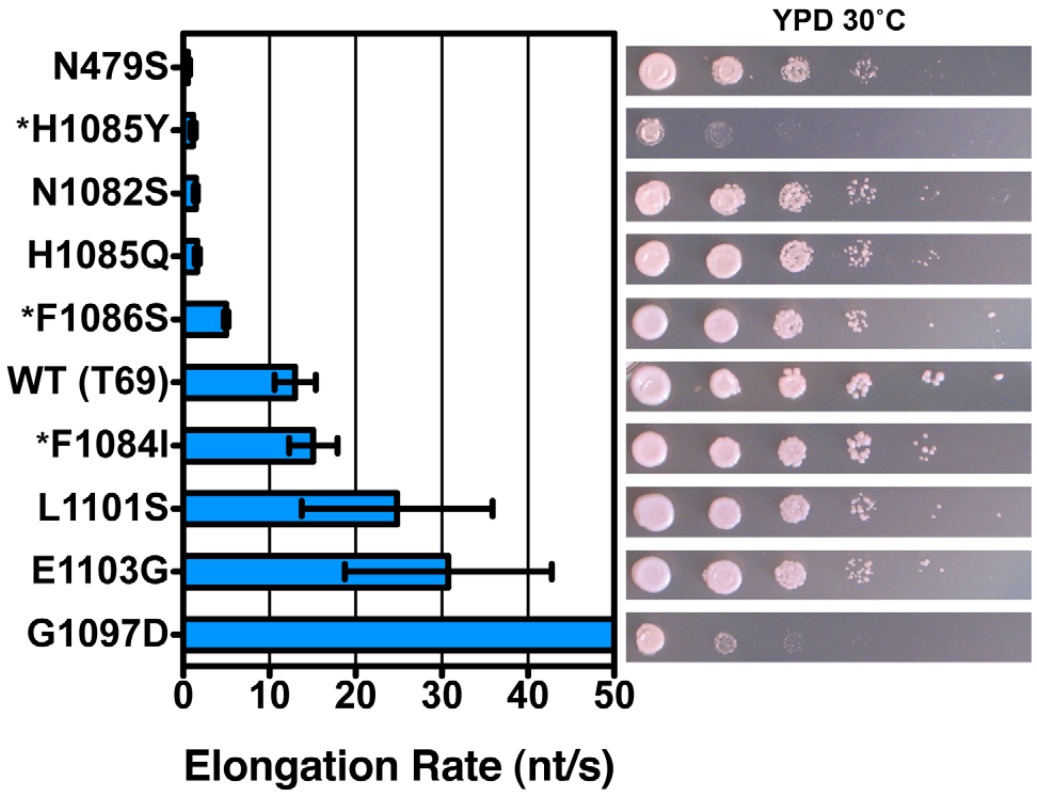


Figure 2-4. Pol II TL substitution mutants show a wide range of elongation rates *in vitro*.

Pol II TL single substitution elongation rates as determined *in vitro* using reconstituted Pol II elongation complexes with synthetic oligonucleotides and an RNA primer (Materials and Methods). Raw data for some mutants (*) are from (Kaplan et al., 2008) and are shown for comparison. Plotted data are maximal elongation rates as determined by non-linear regression of elongation rates determined for different NTP substrate concentrations, and error bars indicate the range of the 95% confidence intervals (Figures 2-5, 2-6). Growth on rich medium of mutants is shown in the right panels for comparison with elongation defects (from Figure 2-3).

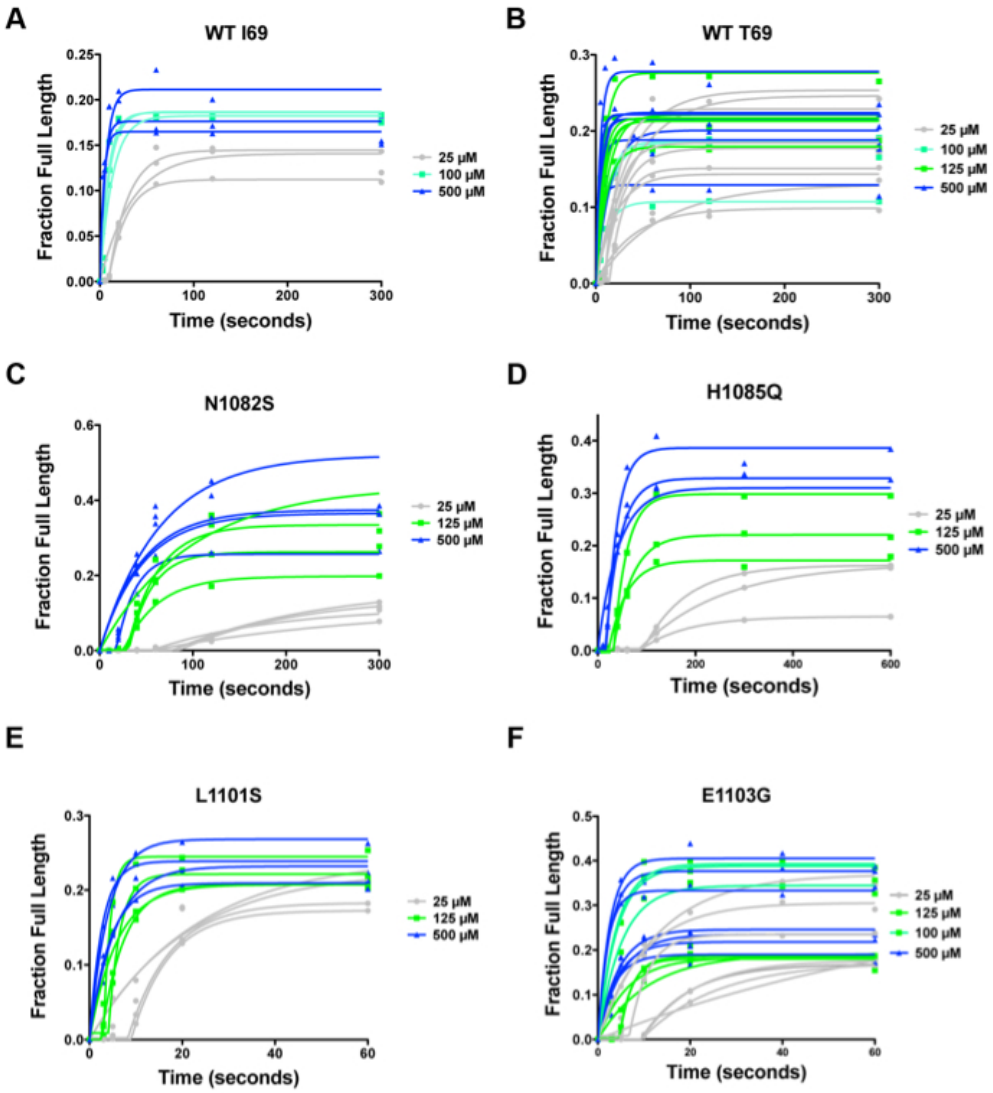


Figure 2-5. Quantification of elongation rates in Pol II mutants.
Run-off transcription as a fraction of total transcription is determined and plotted versus reaction time for specified Pol II mutants. Multiple lines per graph indicate data from time courses generated from separate individual reactions. **A.** WT I69. **B.** WT T69. **C.** N1082S. **D.** H1085Q. **E.** L1101S. **F.** E1103G. **G.** F1086S/E1103G. **H.** H1085Q/E1103G. **I.** H1085Y/E1103G. **J.** N1082A/E1103G. **K.** H1085A/E1103G. **L.** Q1078A/E1103G. **M.** G1097D with standard NTP concentrations. **N.** G1097D with high NTP concentrations ($MgCl_2$ raised to 10 mM from 5 mM).

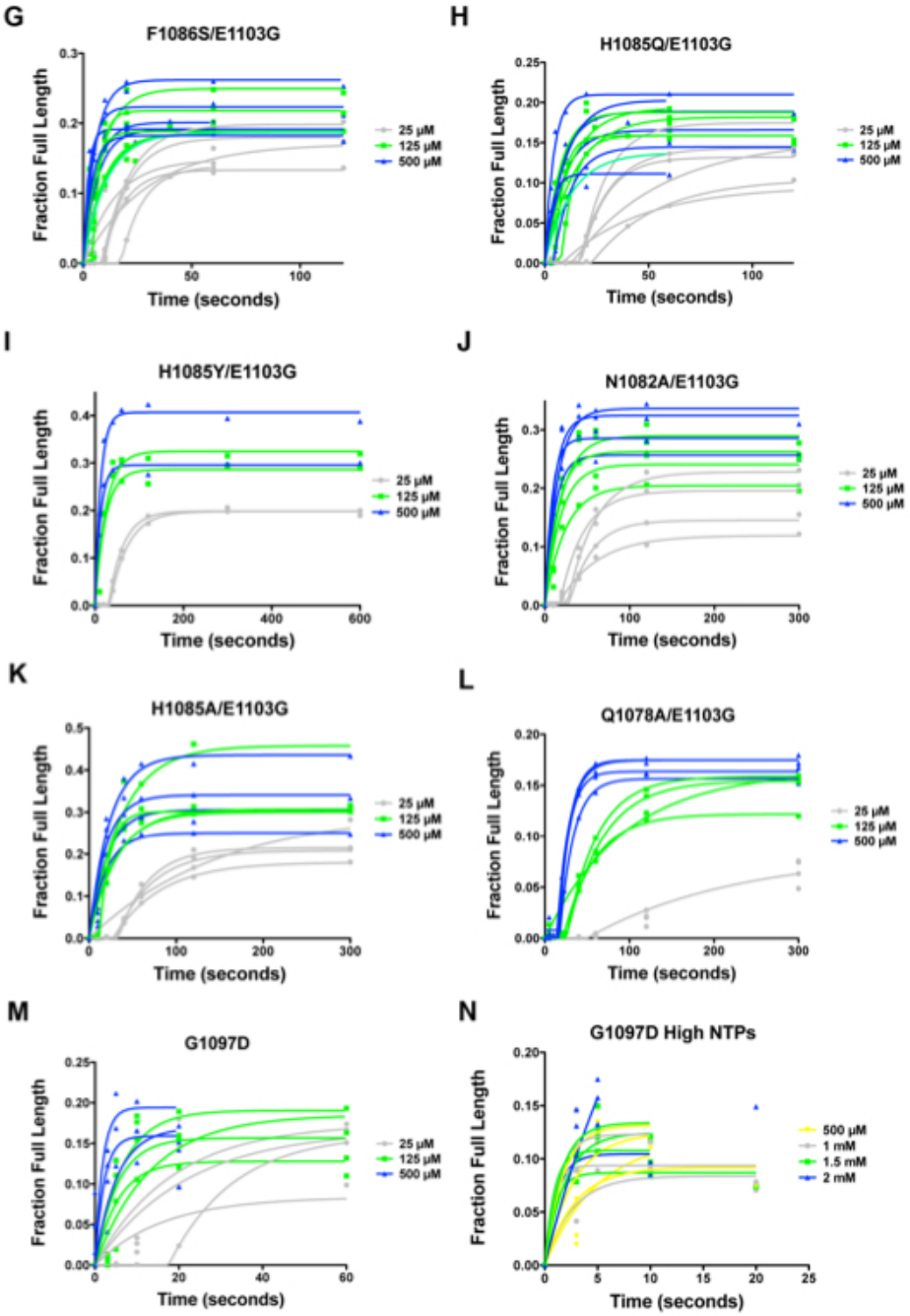


Figure 2-5 Continued.

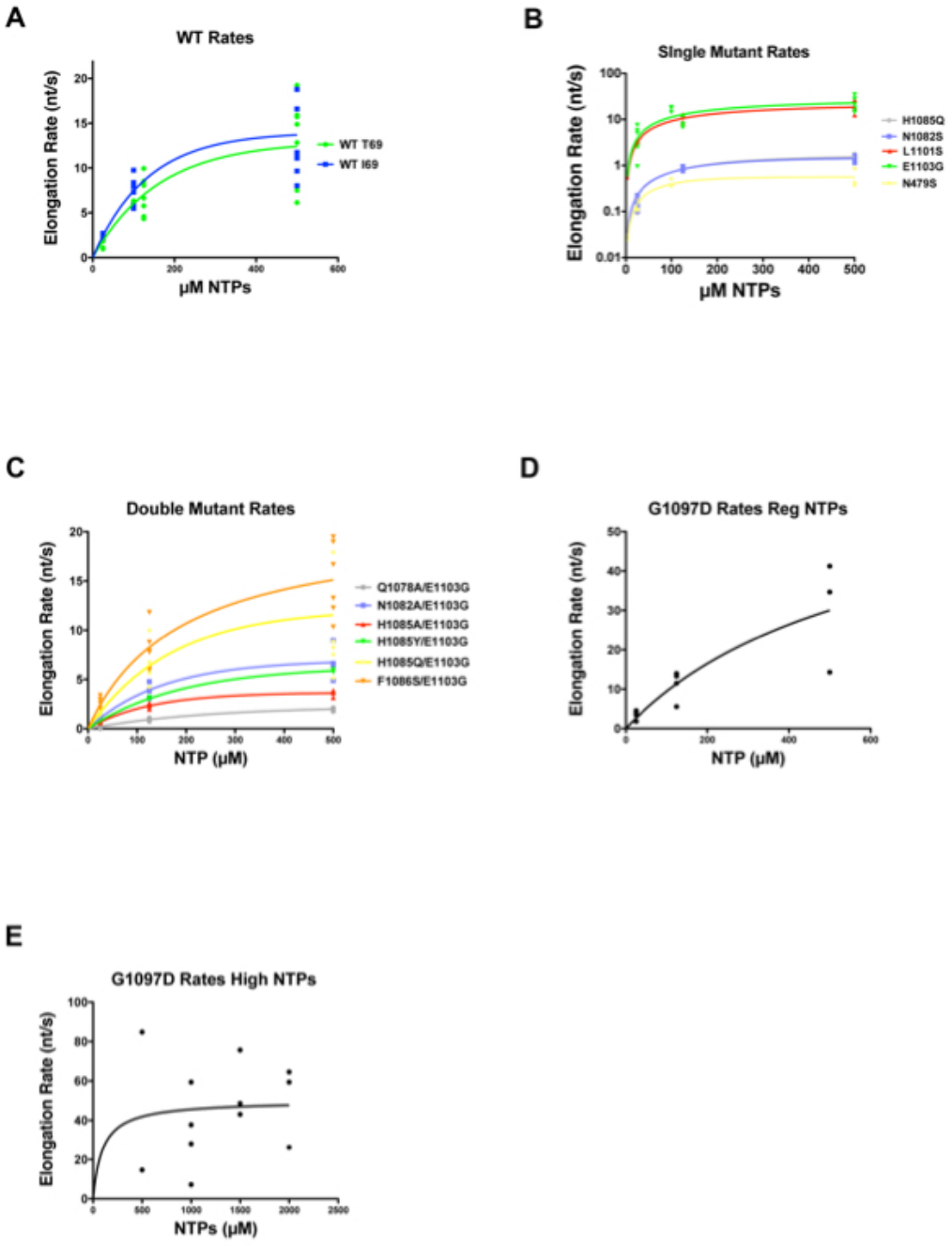


Figure 2-6. Quantification of elongation rates in Pol II mutants (based on 2-5). Rates determined from Figure 2-5 for each NTP concentration are plotted versus NTP concentration and curve fitted by non-linear regression. **A.** WT Pol II enzymes. **B.** Single mutant Pol II enzymes. **C.** Double mutant Pol II enzymes. **D.** G1097D with standard NTP concentrations. **E.** G1097D with high NTP concentrations (MgCl2 raised to 10 mM from 5 mM).

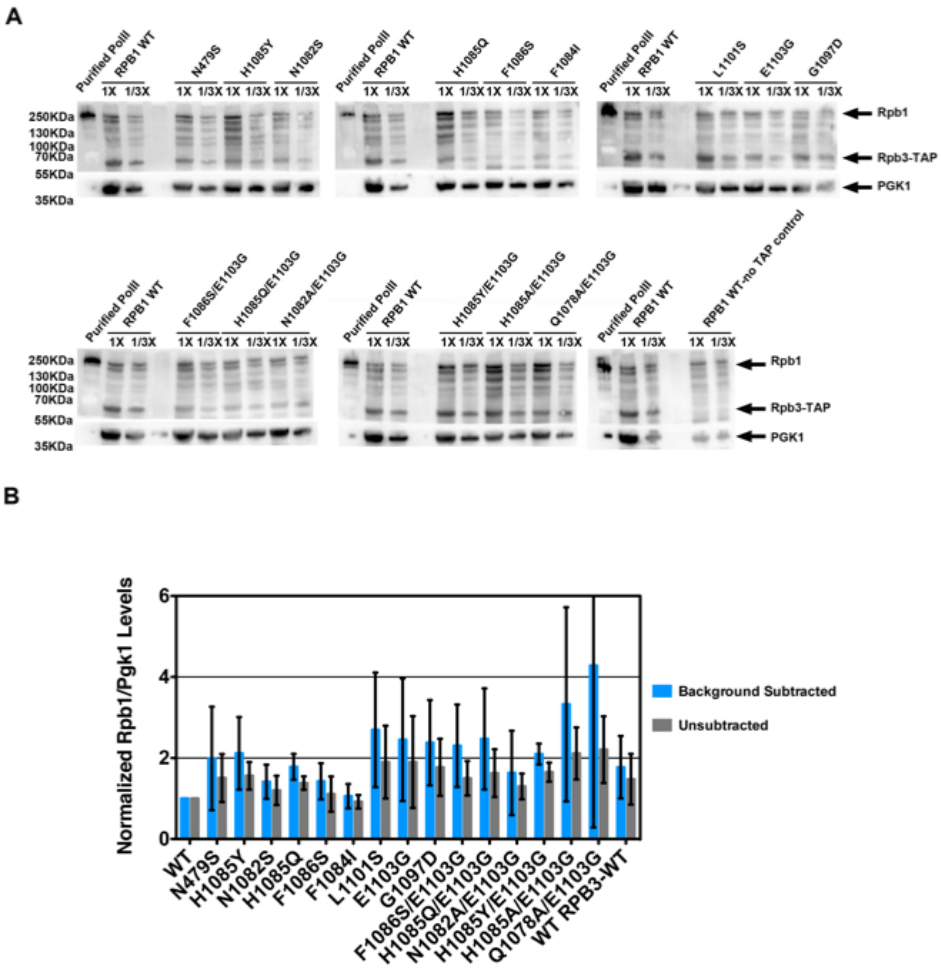


Figure 2-7. Western blotting for Rpb1 and Rpb3-TAP from Pol II variants.
A. Western blotting for Rpb1 expression levels from WT and *rpb1* mutant yeast strains using an antibody to the N-terminus of Rpb1. Because our strains contain Rpb3-TAP, the protein A tag on Rpb3-TAP was also recognized by either the primary or secondary antibodies used. Blotting for Pgk1 was used to confirm equal loading of lanes (bottom panels). Blots shown are representative of three independent experiments. **B.** Quantification of Rpb1/Pgk1 ratio from three independent experiments. Error bars indicate +/- standard deviation of the mean. The anti-Rpb1 antibody has a high background, and subtraction of this background caused standard deviations to generally increase, so both background subtracted and unsubtracted quantifications are shown.

Role Of E1103-T1095 Interactions In Superactivation Of E1103G

For msRNAPs from each kingdom of life, GOF substitutions have been reported in TL residues distal to the NIR and proximal to the C-terminal "hinge" region (Bar-Nahum et al., 2005; Kaplan et al., 2008; Kireeva et al., 2008; Malagon et al., 2006; Tan et al., 2008), which is the portion of the TL that changes conformation from open to the closed NTP-interacting orientation (**Figure 2-1**). There is evidence that the Pol II GOF substitution E1103G promotes a closed TL state (Kireeva et al., 2008). An increased frequency or duration of this state could have consequences for catalysis (increased activity with all substrates and net increase in misincorporation) and translocation (reduced translocation rate likely due to stabilization of closed state after catalysis). The E1103 sidechain interacts with the T1095 sidechain and makes backbone contacts with K1112 in elongation complex crystal structures where the TL is in the open conformation or constrained by other factors (Gnatt et al., 2001; Kettenberger et al., 2004; Westover et al., 2004). Loss of this E1103-T1095 interaction has been posited to explain the effect of E1103G on Pol II transcription (Kireeva et al., 2008). We found that the T1095A substitution was phenotypically indistinguishable from wild type (WT), indicating that loss of T1095-E1103 contacts are not responsible for E1103G growth alteration or, most likely, biochemical superactivity (**Figure 2-3**). T1095R, presenting a much longer sidechain that might disrupt "out conformation" hinge folding, only confers weak MPA^s, Spt^- and Gal^r phenotypes, consistent with C-terminal hinge function or TL dynamics being controlled more strongly by distal residues or other contacts of E1103 (**Figure 2-8**).

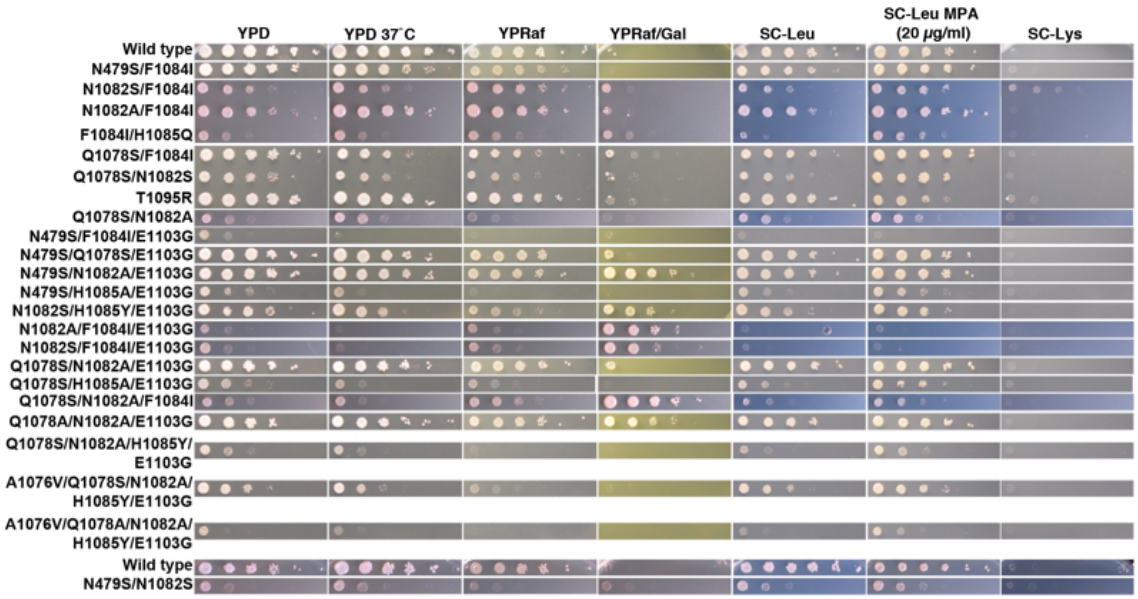


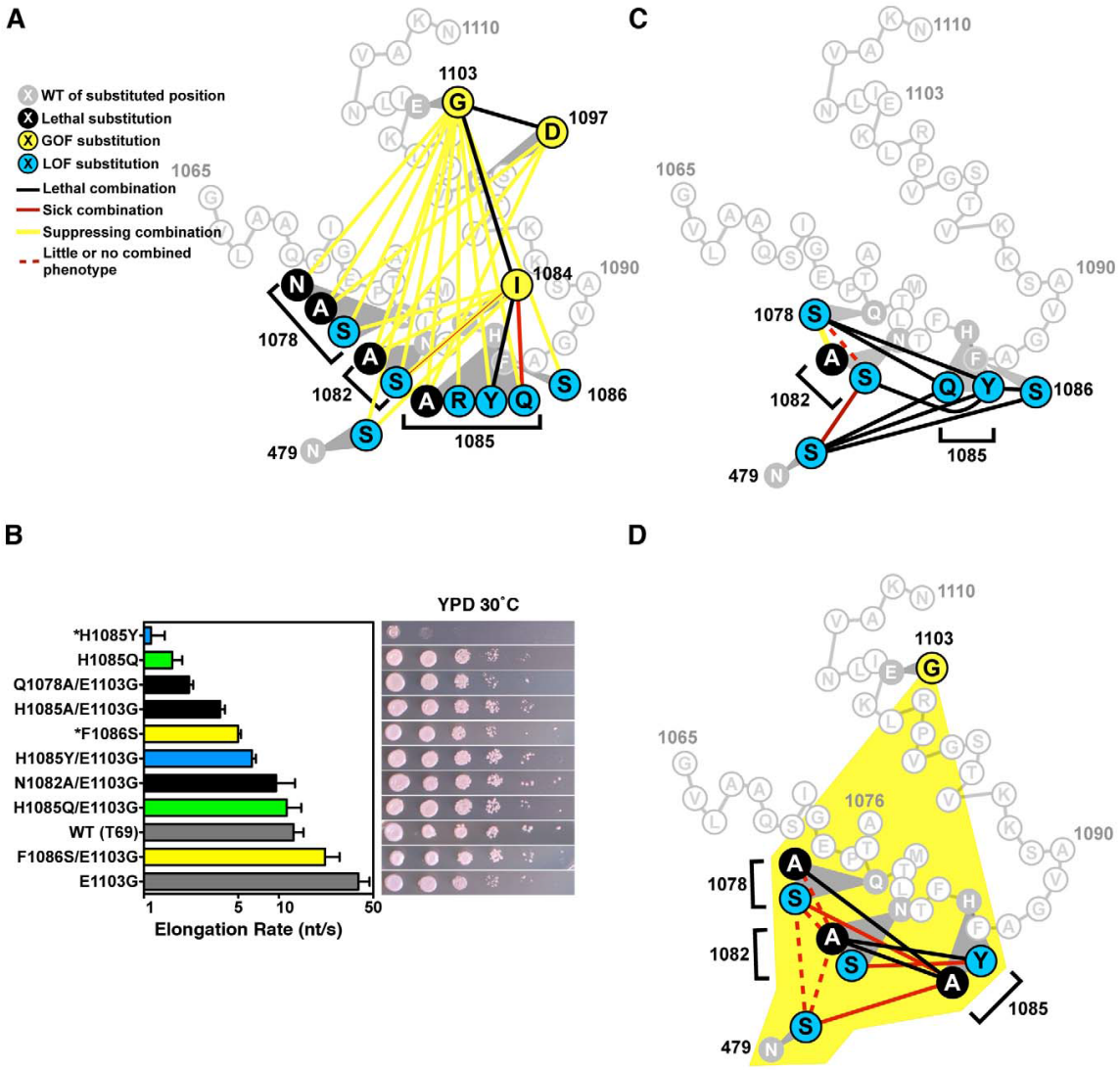
Figure 2-8. Additional Pol II TL single substitution and multiple substitution mutants.

10-fold serial dilutions of saturated cultures of Pol II TL mutant strains plated on different media. Also note that the multiply substituted Q1078S/N1082A/H1085Y/E1103G mutant is viable but confers a strong growth defect. When combined with an A1076V substitution, previously identified as *sit1-8G* (as a double mutant with *rpb1*-N479Y conferring a phenotype reminiscent of the Spt$^-$ phenotype) (Archambault et al., 1998), we observed that A1076V suppressed growth defects of Q1078S/N1082A/H1085Y/E1103G, and that it also confers viability to Q1078A/N1082A/H1085Y/E1103G.

Combinatorial Analyses Of TL Substitutions Reveal Functional Distinctions Between Residues With Similar Single Mutant Behavior

We next wished to probe genetic relationships between TL GOF and LOF substitutions, particularly those between TL hinge residues and TL NIR residues, as these relationships might provide mechanistic insight into TL function (**Figure 2-9** A, **Figure 2-10**, **Figure 2-8**). Surprisingly, by combining TL GOF with TL NIR substitutions, we found that either E1103G or G1097D GOF mutations were able to suppress lethality of many, but not all, inviable (LOF) single substitution NIR mutants (**Figure 2-9** A, **Figure 2-10**). Examination of double mutant plate phenotypes provided more evidence of mutual suppression in E1103G-NIR mutant combinations because Spt^- and MPA^s phenotypes of E1103G were suppressed together with moderate Gal^r phenotypes of individual NIR substitutions (**Figure 2-10**). In contrast, tested pairwise combinations of genetically and biochemically similar GOF substitutions, F1084I, G1097D and E1103G resulted in lethality (**Figure 2-9** A). Additionally, combination of viable LOF NIR-residue substitutions with the GOF substitution E1103G resulted in mutual suppression of conditional plate phenotypes and growth defects (**Figure 2-10**).

Figure 2-9. Combination of TL substitutions alters *in vivo* growth and *in vitro* biochemical phenotypes.
A. GOF-GOF and GOF-LOF genetic interactions among *rpb1* TL and other substitutions. GOF (yellow) and LOF (blue) classifications were determined by biochemical and genetic phenotypes of single substitution mutants, with lethal single substitutions shown in black. Double mutant phenotypes are illustrated as the colored lines connecting colored nodes of particular single mutants. Schematic of TL sequence and orientation is as in Figure 1B. Data were compiled from Figure 2-8, Figure 2-10. N1082S/F1084I shows a complex interaction with exacerbation of growth defects on rich medium but mutual suppression of other conditional plate phenotypes. **B.** *In vitro* elongation rates determined as in Figure 3 for select Pol II TL double mutant enzymes. Single mutants and their relevant E1103G double mutants are indicated by color-coding: F1086 and F1086/E1103G, yellow; H1085Q and H1085Q/E1103G, green; H1085Y and H1085Y/E1103G, blue. Black bars indicate double mutants with E1103G for which the counterpart single substitution mutant is inviable (Q1078A, N1082A, H1085A). Single mutant data are from Figure 3. Raw data for some mutants (*) are from (Kaplan et al., 2008) and are shown for comparison. Error bars indicate the ranges of the 95% confidence intervals. Note that *x*-axis is logarithmic scale. Growth on rich medium of mutants is shown in the right panels for comparison with elongation defects (from Figure 2-3, Figure S2). **C.** Interaction diagram showing phenotypes of pairwise combinations of *rpb1* LOF alleles (legend shown in (A)). **D.** Interaction diagram illustrating genetic interactions between pairs of Rpb1 substitution mutants each in the presence of the E1103G substitution (yellow shading encompassing all relevant residues, legend shown in (A)).


A
WT of substituted position
Lethal substitution
GOF substitution
LOF substitution
Lethal combination
Sick combination
Suppressing combination
Little or no combined phenotype
1065
1090
1110
1103
1097
1084
1078
1082
1086
1085
479
C
1065
1090
1110
1103
1078
1082
1086
1085
479
B
YPD 30˚C
*H1085Y
H1085Q
Q1078A/E1103G
H1085A/E1103G
*F1086S
H1085Y/E1103G
N1082A/E1103G
H1085Q/E1103G
WT (T69)
F1086S/E1103G
E1103G
1
5
10
50
Elongation Rate (nt/s)
D
1065
1090
1110
1103
1076
1078
1082
1085
479

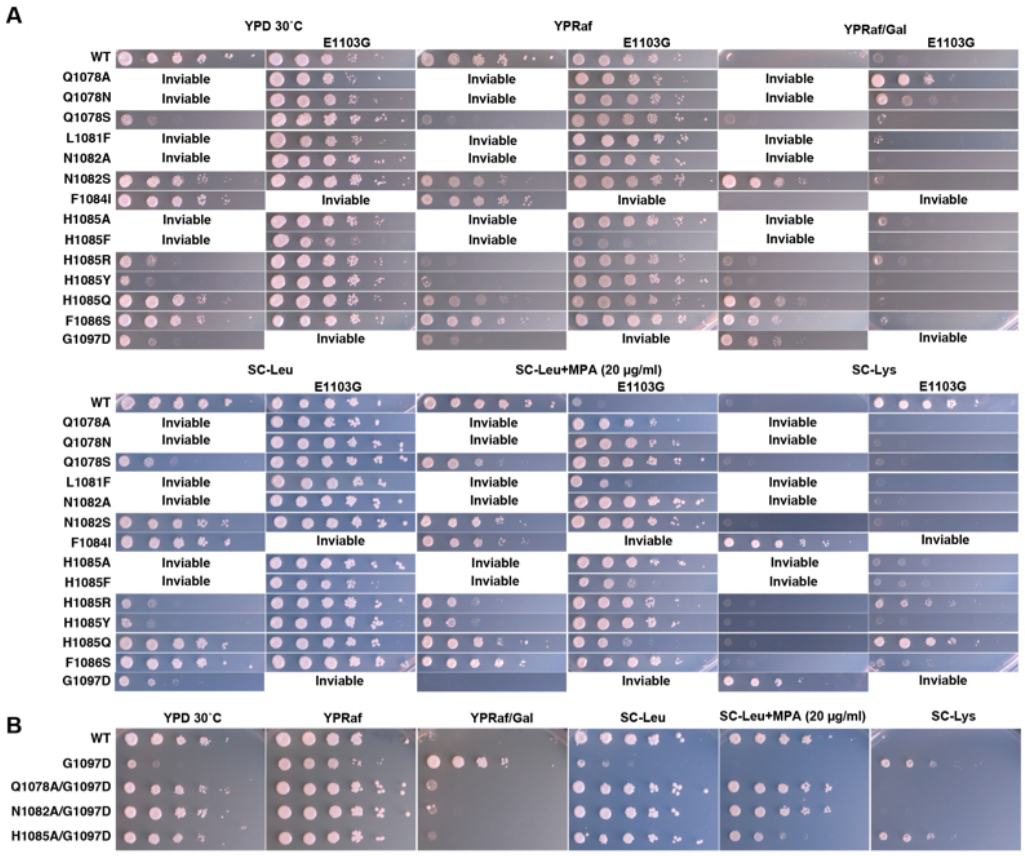


Figure 2-10. Genetic analyses of Pol II TL single substitution mutants combined with E1103G or G1097D substitutions distinguish between different classes of Pol II mutant and show extensive suppressive relationships.

A. 10-fold serial dilutions of saturated cultures for a number of Pol II TL substitutions combined with E1103G plated on different media as in Figure 2-1, with single mutant panels from Figure 2-1 shown for comparison purposes. **B.** 10-fold serial dilutions of saturated cultures for a number of Pol II TL substitutions combined with G1097D are plated on different media as in Figure 2-1.

If observed plate phenotypes directly relate to elongation rate, as suggested by our analysis of single substitution mutations (**Figure 2-4**), the observations above predict that the double mutants' *in vitro* elongation rates might be closer to WT than those of the single substitution mutants. We therefore measured elongation rates of doubly substituted TL mutants to further examine the relationships between GOF and LOF variants (**Figure 2-9** B)(**Figures 2-5** and **2-6**). We found that combination of E1103G and individual viable LOF substitutions F1086S, H1085Q or H1085Y resulted in enzymes with activity intermediate between the relevant singly substituted enzymes. Combination of E1103G individually with the lethal single substitutions Q1078A, N1082A or H1085A resulted in viable strains and Pol II enzymes with activity reduced compared to E1103G alone. The fold enhancement of viable LOF substitutions F1086S, H1085Q and H1085Y by E1103G was roughly similar in each case.

To determine if TL NIR residues function individually or in a coupled fashion, we constructed pairwise combinations using TL NIR LOF alleles and a substitution in the TL-adjacent Rpb1 NTP-interacting residue, N479S (**Figure 2-9** C). We observed lethality when H1085 substitutions were combined with substitutions in other TL NIR residues or with N479S, consistent with H1085 functioning non-redundantly or distinctly from other TL NIR residues or N479 in the Pol II nucleotide addition cycle. Surprisingly, when we examined combinations among N479, Q1078, and N1082 substitutions we did not observe lethality (**Figure 2-9** C). In fact, we observed suppression or epistasis between Q1078 and N1082 substitutions. These observations suggested a functional distinction between N-terminal TL NIR residues (Q1078, N1082) and H1085. Viable

substitutions in these residues maintain sidechains that might still confer some interaction between the TL and NTPs, however alanine substitutions in key residues Q1078, N1082, and H1085 were inviable and could not be directly assessed genetically. Because all LOF TL residues, including these lethal substitutions, were suppressed by E1103G, we examined triple mutants that contained pairwise substitutions in the TL NIR in the background of E1103G (**Figure 2-9** D). These triple mutant combinations revealed similar functional distinctions between two sets of Pol II active site residues, with the set of residues containing N479, Q1078 and N1082 being distinct from H1085. These results may be interpreted in light of structural information and have implications for the mechanism of TL function (see Discussion).

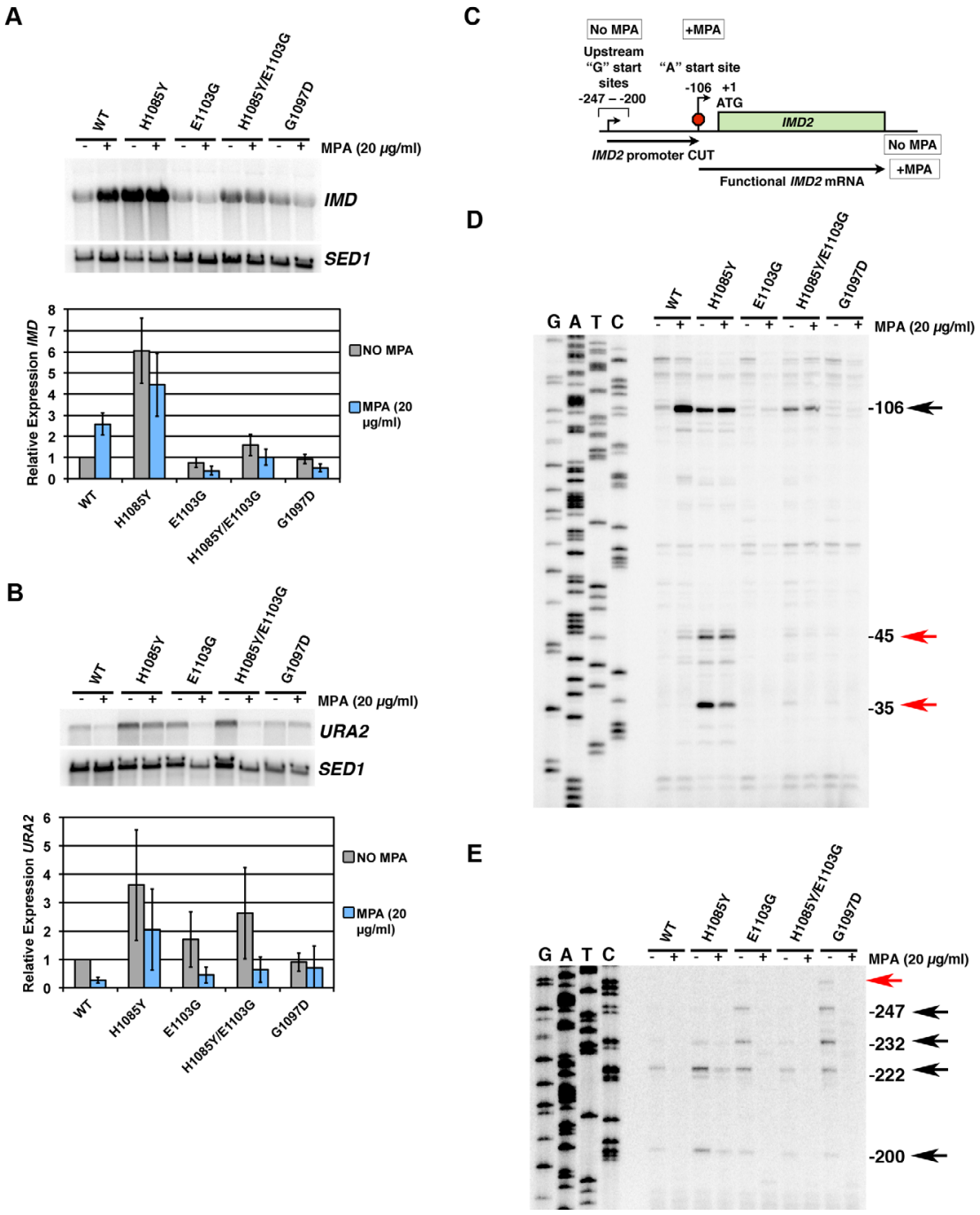
In Vivo Transcription Defects Of Pol II Tl Mutants

To examine our Pol II TL mutants for *in vivo* transcription defects, we analyzed expression of *IMD* genes and *URA2* (**Figure 2-11** A, B), which function in GTP and UTP synthesis, respectively. Expression of one *IMD* gene, *IMD2*, and *URA2*, have been linked to GTP or UTP levels, respectively, possibly directly through alteration of Pol II function by changes in levels of these transcription substrates (Jenks et al., 2008; Kuehner and Brow, 2008; Kwapisz et al., 2008; Thiebaut et al., 2008). Expression of *IMD2* is required for WT yeast to be resistant to the drug MPA (Hyle et al., 2003; Shaw et al., 2001). *IMD2* transcripts are increased in the presence of MPA due to reduced GTP levels that result from inhibition of MPA-sensitive Imd2 homologs, Imd3 and Imd4 (Hyle et al., 2003; Jenks and Reines, 2005). A number of transcription mutants are defective for *IMD2* upregulation upon MPA treatment (Desmoucelles et al., 2002; Riles

et al., 2004; Shaw and Reines, 2000) while some Pol II mutants have been reported to show constitutive expression of *IMD2* (Kuehner and Brow, 2008; Malagon et al., 2006). As *IMD2* has a high level of sequence similarity to the pseudogene *IMD1* and to *IMD3* and *IMD4*, and because of possible cross-hybridization (Hyle et al., 2003), we do not specify that observed changes in our Northern blotting are solely due to *IMD2* changes. However, it has been demonstrated that the majority of MPA-induced *IMD* expression comes from *IMD2* (Hyle et al., 2003).

We observed a constitutive upregulation of *IMD* transcripts in LOF mutant H1085Y and loss of MPA-responsiveness in GOF mutants E1103G and G1097D as well as in the H1085Y/E1103G double mutant (**Figure 2-11** A). E1103G appeared to have reduced expression of *IMD* genes in the absence of MPA treatment, and importantly, E1103G and G1097D showed reduced *IMD* expression upon MPA treatment. These results are consistent with the MPA-sensitivity of E1103G and G1097D and the apparent MPA-resistance of H1085Y (**Figure 2-3**) (Kaplan et al., 2008). They also indicate that defects in *IMD* expression may fully explain the MPA-sensitivity of Pol II TL mutants and need not relate to global elongation defects due to reduced GTP levels, which is the commonly invoked mechanism for MPA-sensitivity. The results, however, are consistent with a model in which *IMD2* regulation occurs directly through Pol II-nucleotide sensing.

Figure 2-11. Pol II TL mutants alter transcription of *IMD2* and *URA2*, genes required for GTP and UTP synthesis, respectively.
A. Northern blotting for different Pol II mutants in the presence or absence of 20 μg/ml MPA (2 hours treatment) for expression of *IMD* gene(s) using *IMD2* DNA probe (top). *SED1* was probed as a loading control. Values for *IMD* normalized to the WT ratio of *IMD/SED1* are shown in the graph and represent the mean relative expression for *IMD* transcript(s) +/- the standard deviation for at least three independent experiments. **B.** Northern blotting for *URA2* expression in the presence or absence of 20 μg/ml MPA (2 hours treatment) (top). *SED1* was probed as a loading control. Values for *URA2* normalized to the WT ratio of *URA2/SED1* are shown in the graph and represent the mean relative expression for *URA2* +/- the standard deviation for at least three independent experiments. **C.** Schematic of *IMD2* gene transcription in the absence or presence of MPA. *IMD2* is not functionally expressed in the absence of MPA or low GTP because upstream transcriptional starts are terminated at a terminator element (stop sign) that can be bypassed by utilization of downstream start sites. **D.** Primer extension analysis for downstream *IMD2* start site usage in Pol II TL mutants in presence or absence of 20 μg/ml MPA (2 hours treatment). Numbers on right indicate position of RNA terminus relative to the A of the *IMD2* ATG codon. Sequence ladder on right is derived from the primer used in (E) and has been cross-referenced with the primer used here. **E.** Primer extension analysis for upstream *IMD2* start site usage in Pol II TL mutants as in (D). Numbers on right indicate position of RNA terminus relative to the A of the *IMD2* ATG codon. Sequence ladder on right is derived from the same primer used. The examples in D and E are representative of at least three independent experiments.


A
WT
H1085Y
E1103G
H1085Y/E1103G
G1097D
MPA (20 μg/ml)
IMD
SED1
Relative Expression IMD
NO MPA
MPA (20 μg/ml)
B
URA2
SED1
Relative Expression URA2
C
No MPA
+MPA
Upstream "G" start sites -247 – -200
"A" start site
-106 +1
ATG
IMD2
IMD2 promoter CUT
Functional IMD2 mRNA
D
G A T C
-106
-45
-35
E
-247
-232
-222
-200

Activation of *URA2* has also been proposed to result from a Pol II-dependent NTP-sensing event because a class of Pol II mutants causes constitutive *URA2* expression (Kwapisz et al., 2008; Thiebaut et al., 2008). *URA2* is normally controlled by a change in start site selection from upstream start sites, which produce transcripts that are prematurely terminated and degraded, to downstream start sites that allow production of full-length mRNA. The Pol II mutants that lead to *URA2* expression are located in the "switch 1" region of Rpb1 (Cramer et al., 2001), near a residue linked to start site selection (Berroteran et al., 1994). We reasoned that if *URA2* were directly responsive to Pol II-sensing of UTP levels, then a Pol II TL LOF mutant should also show constitutive upregulation of *URA2*, as a reduction of Pol II activity through TL defects should mimic reduction in substrate levels. We found that H1085Y caused *URA2* to be upregulated (**Figure 2-11** B). GOF TL mutants E1103G and G1097D were closer to WT for *URA2* expression, while H1085Y/E1103G showed an increase in *URA2* expression. Furthermore, *URA2* was not responsive to MPA treatment, an expected result if *URA2* were specifically responsive to UTP depletion as opposed to GTP depletion.

IMD2, like *URA2*, is regulated by a shift in start sites from upstream starts that lead to premature termination, to downstream starts upon nucleotide limitation (Jenks et al., 2008; Kuehner and Brow, 2008). The start site changes in *IMD2* are proposed to relate to concentration of initiating NTP, with upstream starts initiating with GTP, whereas under GTP-limiting conditions, the major downstream start initiates with ATP (**Figure 2-11** C). We examined this phenomenon in Pol II TL mutants by primer extension (**Figure 2-11** D, E). Examination of downstream start sites at *IMD2* showed

that the major ATP start site at −106 from the *IMD2* start codon is induced by MPA treatment in WT cells (**Figure 2-11** D) (Steinmetz et al., 2006). TL LOF H1085Y showed constitutive expression from this start regardless of MPA treatment, and, interestingly, apparent usage of downstream cryptic starts at approximately -35 and -45. Although mRNA processing in theory could generate these shorter new transcripts, it is unlikely as processed transcripts lacking a 5′-cap should be fairly unstable. In contrast, for TL GOF mutants E1103G and G1097D little or no usage of the −106 start was observed, regardless of MPA treatment, consistent with MPA-sensitivity of these strains and Northern blotting showing no MPA induction of *IMD* transcripts. In the H1085Y/E1103G mutant that showed mutual suppression of respective single mutant phenotypes but reduced activity from WT *in vitro*, we observed constitutive usage of the −106 start, but at apparent lower levels than H1085Y. In addition, we observed MPA-dependent loss of upstream start site usage in this set of mutants, suggesting maintained responsiveness to MPA-effects for these starts, even though H1085Y showed constitutive use of downstream starts. In contrast to the observed H1085Y usage of new -35 and -45 downstream start sites (**Figure 2-11** D), we observed increased E1103G and G1097D usage of upstream starts relative to WT (**Figure 2-11** E), suggesting altered start site selection in Pol II TL mutants.

Activity-Dependent Control Of Pol II Start Site Selection In Vivo

A number of Pol II subunit and general transcription factor (GTF) mutants have been found to alter start site selection in *S. cerevisiae* (Berroteran et al., 1994; Eichner et al., 2010; Freire-Picos et al., 2005; Furter-Graves et al., 1994; Ghazy et al., 2004; Hull et

al., 1995; Majovski et al., 2005; Pappas and Hampsey, 2000; Pinto et al., 1992; Sun and Hampsey, 1995, 1996), but it is unclear how Pol II activity might relate to these start site defects. For example, deletion of the Rpb9 subunit of Pol II confers many similar activity defects as Pol II TL GOF substitutions such as increased rate of elongation and misincorporation (Koyama et al., 2007; Nesser et al., 2006; Walmacq et al., 2009) and is defective for start site switching at *IMD2* in response to MPA (Jenks et al., 2008). Additionally, *rpb9*Δ has long been known to exhibit upstream shifts in start site selection (Furter-Graves et al., 1994; Hull et al., 1995; Sun and Hampsey, 1996), while many mutations throughout Pol II can cause downstream shifts in start site selection (references above). Models for Pol II residue function in start site selection have previously focused on physical or functional interactions between these residues and GTFs. We propose that the Pol II TL primarily or exclusively functions in controlling the Pol II nucleotide addition cycle (substrate selection, catalysis and translocation), which, for this work, we define as Pol II activity. Therefore we used our TL mutants to interrogate the relationship between Pol II biochemical activity and start site selection *in vivo*, following our observations of downstream start site usage at *IMD2* in TL LOF H1085Y and apparent upstream shifts in TL GOF E1103G and G1097D mutants. We examined start site selection at a number of other genes by primer extension analysis (**Figure 2-12**).

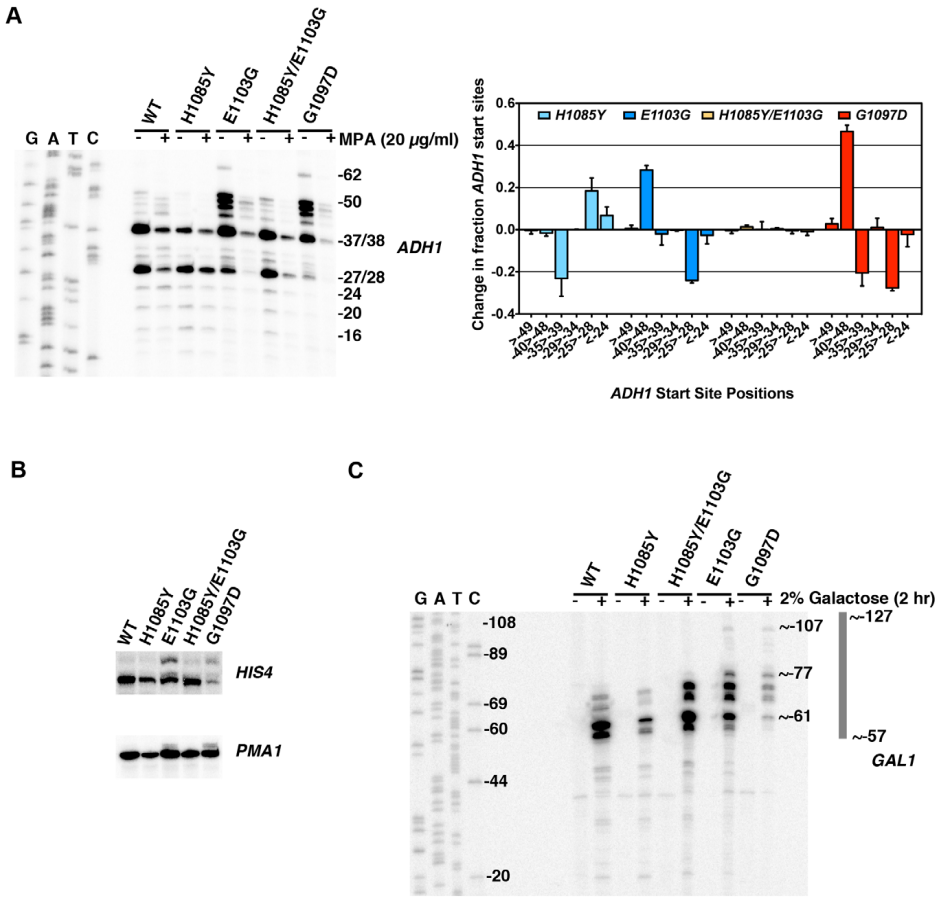


Figure 2-12. Pol II TL contributes to start site selection at a number of genes. **A.** Primer extension analysis of RNA 5′-ends at *ADH1* in Pol II TL mutants. Sequence ladder at left is derived from primer used in Figure 5E but has been cross-referenced with *ADH1* primer used for primer extension here. Numbers indicate positions of putative start sites in relation to the *ADH1* ATG (A is +1). Right panel quantifies start site usage at *ADH1* as determined by primer extension with a radiolabeled oligo. Signals in each lane were divided into bins based on positions of start sites relative to *ADH1* ATG sequence (A is +1) and normalized to total signal per lane. WT start site fractions were subtracted from mutant start site fractions for particular regions to determine the relative alteration in start site distribution in Pol II mutant strains. A negative value indicates that the mutant has relatively lower usage for that particular group of start sites and a positive value indicates a relatively higher usage for that particular group of start sites. Values shown are the average of at least four independent experiments +/- standard deviation. **B.** Primer extension analysis of RNA 5′-ends at *HIS4* and *PMA1* in Pol II TL mutants. **C.** Primer extension analysis of RNA 5′-ends at *GAL1* in Pol II TL mutants. Cells are grown in medium lacking raffinose and then *GAL1* is induced by addition of galactose to 2% for two hours. Sequence ladder at left is derived from the same primer used for primer extension. Numbers indicate positions of putative start sites in relation to the *GAL1* ATG (A is +1).

We found that TL LOF H1085Y caused downstream shift in start site distribution at *ADH1*, similar to reported defects of a number of other Pol II mutants and GTF mutants (**Figure 2-12** A). In contrast, TL GOF E1103G and G1097D substitutions confer upstream defects similar to those that occur in *rpb9*Δ (references above), a number of TFIIF mutants (Eichner et al., 2010; Freire-Picos et al., 2005; Ghazy et al., 2004; Majovski et al., 2005; Sun and Hampsey, 1995), and novel insertion mutants within TFIIB (Zhang et al., 2002). As with other *in vivo* and *in vitro* phenotypes, combination of H1085Y and E1103G leads to mutual suppression, extending the correlation between observed phenotypes and *in vitro* activity. For transcripts from *HIS4* and *PMA1*, genes known to be sensitive to start site alterations in mutant strains (Berroteran et al., 1994; Hull et al., 1995), we also observed upstream shifts in E1103G and G1097D-bearing strains (**Figure 2-12** B). E1103G-dependent start site changes were suppressed by H1085Y for these genes as well. It is not obvious how the alteration of TL function would impact start site selection so strongly, as we expect TL function to be downstream of initiation events important for start site selection, such as pre-initiation complex formation and promoter melting. The strong upstream starts observed for GOF mutants led us to predict that these positions should already be accessible to Pol II in WT cells, because TL function in controlling Pol II activity should be downstream of promoter melting. Using permanganate footprinting to identify regions of single-stranded DNA in the *GAL1* and *GAL10* promoters, Giardina and Lis observed strong permanganate reactivity at greater than 60 basepairs (bp) upstream of the major transcriptional start sites, indicating promoter melting far upstream (Giardina and Lis,

1993). We reasoned that *GAL1* should show far upstream start site shifts in TL GOF mutants within this melted region though it does not normally support productive start sites. Indeed, we observed new start site usage more than 40 bp upstream of the major *GAL1* start sites in TL GOF mutants, and a slight shift in starts sites downstream in TL LOF H1085Y (**Figure 2-12** C, **Figure 2-13**).

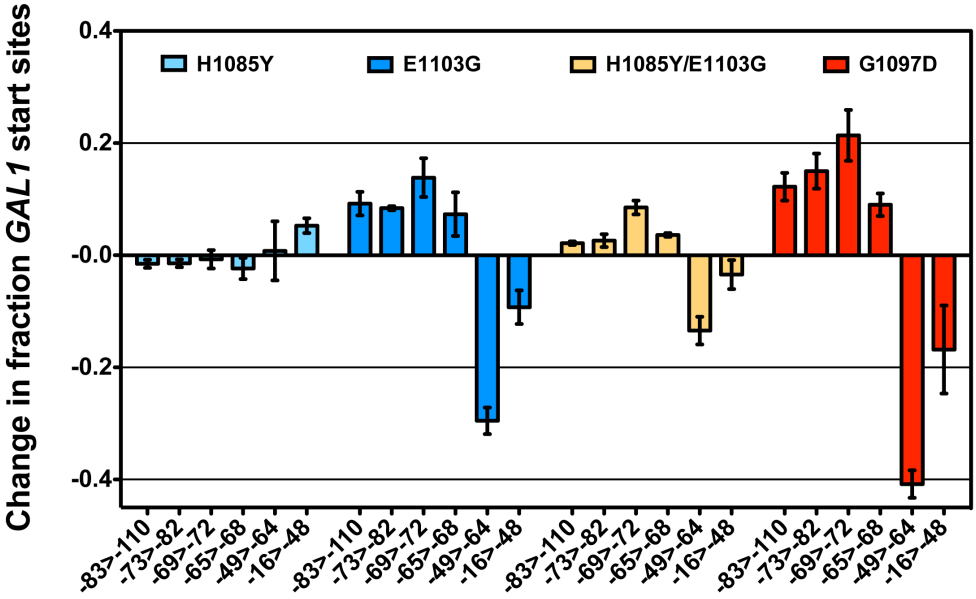


Figure 2-13. Quantification of start site distribution at *GAL1* in Pol II mutants. Start site usage at *GAL1* as determined by primer extension with radiolabeled oligo was divided into bins based on radioactive signal at positions of start sites relative to *GAL1* ATG sequence (A is +1) and normalized to total signal per lane. WT start site fractions were subtracted from mutant start site fractions for particular regions to determine the relative alteration in start site distribution in Pol II mutant strains. A negative value indicates that the mutant has relatively lower usage for that particular group of start sites and a positive value indicates a relatively higher usage for that particular group of start sites. Values shown are the average of three independent experiments +/- standard deviation.

Discussion

The contributions of TL residues to the Pol II nucleotide addition cycle ("activity" for the purposes of this discussion) are critical for transcription *in vitro* and *in vivo*. The catalytic contribution of the TL has been proposed to function through minimization of NTP dynamics within the Pol II active site, which in turn could increase the probability of catalysis (Huang et al., 2010). Our biochemical experiments coupled with extensive genetic analyses provide insight into the function of TL residues. We calculated an approximate 6-fold suppression effect on elongation rate conferred by the E1103G GOF substitution when it was combined with LOF substitutions. We used this fold enhancement to infer the putative defects of inviable TL LOF substitution mutants for which we were unable to measure elongation rate directly. Assuming an approximate 6-fold increase in rate contributed by E1103G to the values of E1103G-suppressed inviable substitution elongation rates, we infer 7-30 fold defects for N1082A, H1085A and Q1078A single substitutions (**Figure 2-9** B). Examination of analogous substitutions in *E. coli* or *T. thermophilus* RNAPs has revealed large differences in the contribution of these residues to catalysis with NTP substrates (4 and 6 fold for *Eco* R933A$^{Sce\text{N1082A}}$ and H936A$^{Sce\text{H1085A}}$, respectively (Zhang et al., 2010) vs. 50 and 100 fold for *Tth* R1239A$^{Sce\text{N1082A}}$ and H1242A$^{Sce\text{H1085A}}$, respectively (Yuzenkova et al., 2010)). Our results indicate that *S. cerevisiae* Pol II TL residues have contributions intermediate to TL residues from different bacteria and suggest that contributions of TL residues to activity cover a range of values in homologous enzymes.

Our data also support a functional relationship between regions adjacent to TL hinges, where the TL has been observed to deviate from its fully folded, NTP-bound conformation in crystal structures. These regions are the location of most of the strong Spt^- and MPA^S TL substitutions, as well as the strongest GOF alleles tested. These results underscore the idea that the TL is finely balanced between functional states or conformations, and that the hinges are sensitive to changes that in some cases can promote Pol II activity. These changes likely occur at the expense of fidelity, as has been shown for E1103G (Kaplan et al., 2008; Kireeva et al., 2008). Alteration of distal TL residues (*e.g.* E1103, G1097D) conferred broad suppression on a number of TL NIR substitutions. Thus, it appears that loss of NIR function can be compensated by alteration in trigger loop dynamics. Importantly, functions of TL NIR residues are not completely bypassed, as activities of tested double mutants are intermediate between the single mutants, and not all NIR mutants are suppressible.

Functional distinctions between TL residues and genetic epistasis between some residue combinations have been hinted at in the case of residues analogous to Q1078 and N1082 in *T. thermophilus* experiments (Yuzenkova et al., 2010), but our combinatorial genetic analyses reveal this more clearly (**Figure 2-9**). Examination of different Pol II crystal structures shows the TL in a number of states between partially and stably folded in the presence or absence of matched NTP or NTP analog (Cheung et al., 2011; Wang et al., 2006; Westover et al., 2004)(**Figure 2-14**).

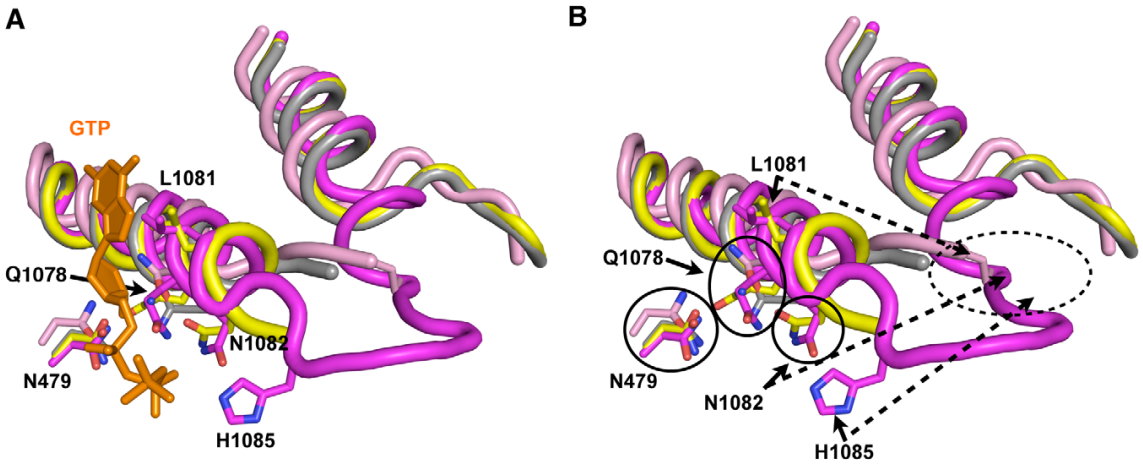


Figure 2-14. Model for stepwise function of TL residues in contributing to TL folding and function.

A. Structural cartoon showing overlay of TLs and Rpb1 N479 from a number of Pol II elongation complex/initial transcribing complex crystal structures with TLs in different states of folding/NTP interaction. Matched GTP nucleotide (orange) and TL and N479 (magenta) are from PDP 2E2H (Wang et al., 2006), which shows TL in a closed conformation. Partially folded TL and N479 (yellow) from PDB 4A3F (Cheung et al., 2011), which contains a matched NTP analog (not shown) and short RNA (not shown). Related structure to PDB 4A3F from PDB 4A3D (Cheung et al., 2011), but without matched NTP analog is shown in gray. Partially folded TL from PDB 1R9T, containing a mismatched NTP (not shown) is shown in pink. **B.** Same figure as in (A) but with GTP omitted. Certain residues are relatively closely positioned in all structures overlaid (N479, Q1078), while other residues are observed stabilized in partially folded structures directly adjacent to the active site (L1081, N1082). H1085 is observed positioned adjacent to the active site only upon complete folding/stabilization of the TL. Dashed oval indicates likely location of TL NIR residues when in out conformation where the TL is either mobile or unfolded. This figure was created with Pymol (Schrodinger, 2010).

These structures suggest a TL folding pathway wherein N-terminal TL residues that are proximal to the N-terminal TL hinge (primarily Q1078), along with other active site residues (for example, Rpb1 R446 and N479), are positioned essentially at all times for NTP interaction. Biochemical experiments indicate that these residues function in ribose 2′- and 3′-hydroxyl (OH) recognition, with the 2′-OH of course critical for *selection* of NTPs over 2′-dNTPs (Svetlov et al., 2004; Wang et al., 2006; Yuzenkova et al., 2010). Upon matched NTP basepairing with the DNA template, NTP interactions with R446, N479 and Q1078 may stabilize or begin to promote TL-folding in an N-terminal to C-terminal fashion, where L1081 and N1082 next come into position via N-terminal TL hinge movement. Supporting such a step, in some Pol II crystal structures, these residues are observed adjacent to a matched NTP basepairing with the template, but without complete folding or stabilization of the TL (**Figure 2-14**). Our genetic analyses suggest that Q1078 and N1082 (along with N479) are functionally interdependent (see also Text S1). Following initial hinge movement and TL folding, H1085 and the rest of the TL loop tip must fold into the closed conformation, and genetic analyses of H1085 substitutions suggest this step is distinguishable from functions of other TL NIR residues. It is important to note that we have not been able to evaluate contributions of residues L1081 and R446 in our genetic system because all tested substitutions in these residues were inviable and these substitutions were generally not suppressed by E1103G.

The Relationship Between MPA-Sensitivity And Pol II Elongation Defects

In vivo it has been observed that some Pol II mutants constitutively express *IMD2* in the absence of drug-induced stimulation (Desmoucelles et al., 2002; Malagon et al., 2006). It is unclear whether this constitutive expression represents a mutant-induced alteration in Pol II start site specificity or a generic reduction of Pol II activity that mimics a drug-induced nucleotide depleted state, which itself leads to altered start site selection. We show here that an MPA-resistant LOF TL allele caused constitutive *IMD2* expression, and in other work we found that this expression is likely a generic hallmark for reduced Pol II activity (Braberg et al, in preparation), and not necessarily due to a specific defect in initiation as previously proposed (Kuehner and Brow, 2008). These results support published models that *IMD2* regulation is directly responsive to reduction in nucleotide levels as reduction or increase in Pol II catalytic rates have opposing effects on the ability of *IMD2* to be induced. Our observations that reduced Pol II activity does not correlate with sensitivity to MPA suggest that generic Pol II catalytic defects do not confer sensitivity to reduced nucleotide levels. Given the fact that LOF TL mutants are highly likely to confer elongation defects *in vivo*, the absence of MPA-sensitivity raises the question of whether reduction of GTP levels sensitizes cells to generic elongation defects caused by *any* transcription factor proposed to be acting through Pol II elongation. MPA sensitivity of any transcription factor mutant should not be used to infer presence or absence of an elongation defect because actual nucleotide levels in different mutant strains may vastly differ due to *IMD2* transcriptional effects.

Model For Start Site Selection In Budding Yeast

The observation that TL mutants show polar effects on start site selection *in vivo* correlating with *in vitro* catalytic defects suggests a model for how Pol II activity may directly influence initiation. Start site selection in *S. cerevisiae* is not restricted to a short window a defined distance from the TATA box, but instead can take place over a longer region downstream (Struhl, 1989). It has been proposed that Pol II scans for sequences permissive for productive elongation (Giardina and Lis, 1993). We propose that our mutants primarily affect the nucleotide addition cycle and not some other step in initiation, therefore we expect promoter melting is already occurring in regions where start sites are not normally utilized in WT cells, and that scanning occurs subsequent to template melting, as has been proposed for *GAL* genes (Giardina and Lis, 1993). *Continuous* RNA synthesis during the scanning process does not appear to be required for scanning, as experiments show that chain-terminating nucleotides do not abrogate scanning *in vitro* (Khaperskyy et al., 2008; Lue et al., 1989). In light of these results, models for start site scanning in yeast that do not invoke transcription have been proposed (*i.e.* (Miller and Hahn, 2006), also reviewed in (Corden, 2008). We show that substitutions within the Pol II TL that have increased activity *in vitro* shift start sites upstream at *ADH1*, *HIS3*, *PMA1* and *GAL1*, and those with reduced *in vitro* activity shift start sites downstream at *ADH1* and *GAL1* (**Figure 2-12** and **Figure 2-13**). To explain these data, we invoke a new transcription-assisted model for start site scanning (**Figure 2-15**).

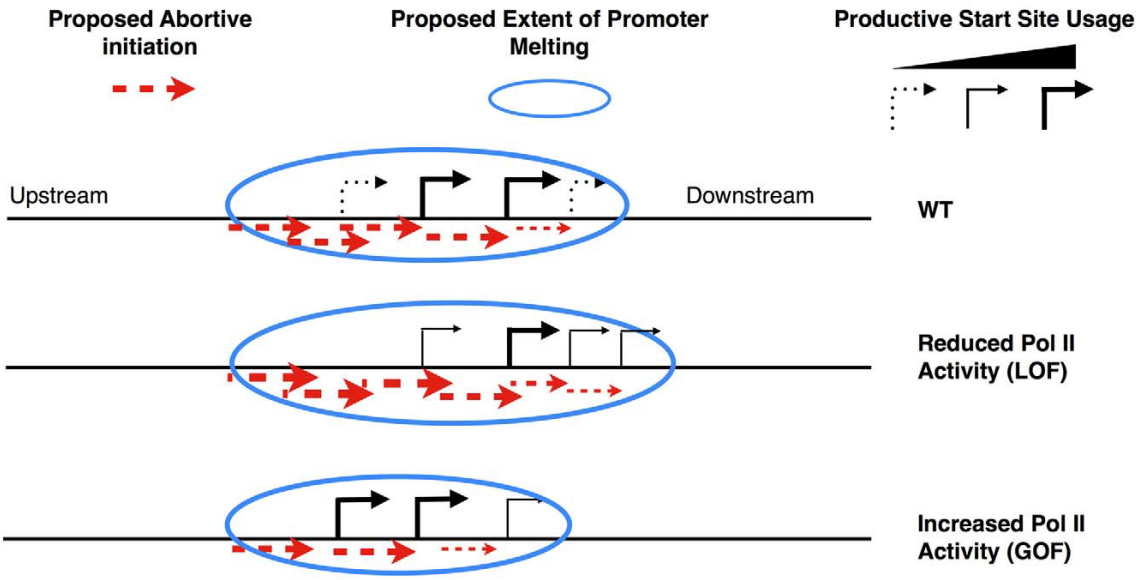


Figure 2-15. Model for transcription-assisted start site scanning through abortive initiation.

Cartoon showing predicted start site distribution for an *ADH1*-like gene in WT cells (top) or alteration in distribution in the presence of LOF (middle) or GOF (bottom) Pol II TL mutants.

Inefficiently or apparently unused start sites may allow undetected *abortive* but not productive initiation. Such abortive transcription may support Pol II downstream translocation, providing an explanation how Pol II catalytic mutants alter start site selection in a polar fashion, which is indicative of altered scanning as described in the flux model proposed by the Brow lab (Kuehner and Brow, 2006). In the flux model, Pol II scanning from upstream to downstream will utilize starts in a polar fashion, and a decrease in usage of an upstream start will lead to increased usage of a downstream start. In our model, upstream shifts in start sites are due to increased Pol II catalytic activity, which in turn increases the probability of normally abortive initiation events becoming productive at the expense of downstream events. Conversely, downstream shifts in start sites are due to decreased Pol II activity, which increases the probability that abortive initiation will occur at upstream positions at the expense of productive initiation from those positions. Competition between abortive and productive initiation may occur at key transitions for early elongation complex stability (Kugel and Goodrich, 2002), or by modulating stability of RNA:DNA hybrid in early transcription complexes as suggested by recent structural studies from the Kornberg lab (Liu et al., 2011) and the Cramer lab (Cheung et al., 2011). These new structures demonstrated a role for a matched NTP substrate in the Pol II active site in stabilization of short nascent RNAs. As GOF Pol II TL alleles have been implicated in extending the duration of substrate-bound complexes (Kireeva et al., 2008), these GOF alleles may also stabilize early elongation complexes in addition to increasing elongation rate. Any increase in the probability of next

nucleotide incorporation at the expense of transcript release would lead to a net *increase* in productive elongation and a net *decrease* in downstream scanning by our model.

Previously described start site-altering mutants in Pol II general factors and other Pol II subunits have activities that are entirely consistent with this model. *rpb9*Δ shows upstream shifts in start sites at *ADH1* (and other genes) (Hull et al., 1995; Sun et al., 1996), inability to shift to downstream *IMD2* starts upon MPA treatment, and causes Pol II to be superactive *in vitro* (Walmacq et al., 2009), similarly to superactive TL mutants E1103G and G1097D. An allele of the TFIIF subunit-encoding *TFG1* gene causes upstream start site shifts *in vivo* while increasing efficiency of Pol II initiation in an abortive initiation assay, consistent with this allele conferring an increase in Pol II activity (Khaperskyy et al., 2008). This increase in Pol II activity may be due to a gain of function in TFIIF or a loss of a negative function, under our activity-dependent framework for interpreting start site defects. Consistent with the latter, a recent *in vitro* study shows that TFIIF can inhibit initiation at specific template positions on a modified *HIS4* template (Fishburn and Hahn, 2012). On some of the templates used in this study, TFIIF inhibits transcription initiation from upstream positions, consistent with polarity of most start site shifting mutants of TFIIF subunits *in vivo*. A number of *sua7* (TFIIB) alleles show downstream shifts in start site selection (Pinto et al., 1992; Pinto et al., 1994; Wu et al., 1999), and in the case of the highly studied E62K allele, this mutant causes a decrease in Pol II transcription efficiency *in vitro* (Cho and Buratowski, 1999). We provide a framework for interpreting the relationship between start site selection changes and the nature of alteration of the initiation process. Increasing or decreasing efficiency

of productive initiation from particular nucleotide positions may relate to upstream or downstream shifts in start site selection, which appear to correlate with increases or decreases in Pol II activity, respectively. While start site defects strongly correlate with Pol II activity, their contribution to the growth defects of Pol II TL mutants remains to be determined.

Materials and methods

Genetic Analyses

To examine mutant polymerase alleles encoding substitutions in the TL and elsewhere in Rpb1, we perform a plasmid shuffle with *rpb1 CEN LEU2* plasmids by assaying ability of *rpb1*Δ cells to grow without an *RPB1 CEN URA3* plasmid. This is accomplished by treatment of cells with the drug 5-fluoroorotic acid (5-FOA, Gold Biotechnology), which is toxic to $Ura3^{+}$ cells (Boeke et al., 1987), thus selecting cells containing solely an *rpb1 CEN LEU2* plasmid so we might assess the ability of mutant plasmids to complement the essential function of *RPB1*. During the course of our experiments, we identified a previously unreported polymorphism in our *RPB1* plasmid encoding substitution of isoleucine at position 69 for the threonine reported for this position in the Yeast Genome Database (T69I). This substitution was tracked as far back as the pRP112 subclone derived from an original *RPO21* genomic clone from the Young lab (Nonet et al., 1987). Our analyses indicate that this substitution is phenotypically inert (**Figure 2-16**), yet all genetic data shown are for plasmids with the Ile69 corrected to threonine (I69T).

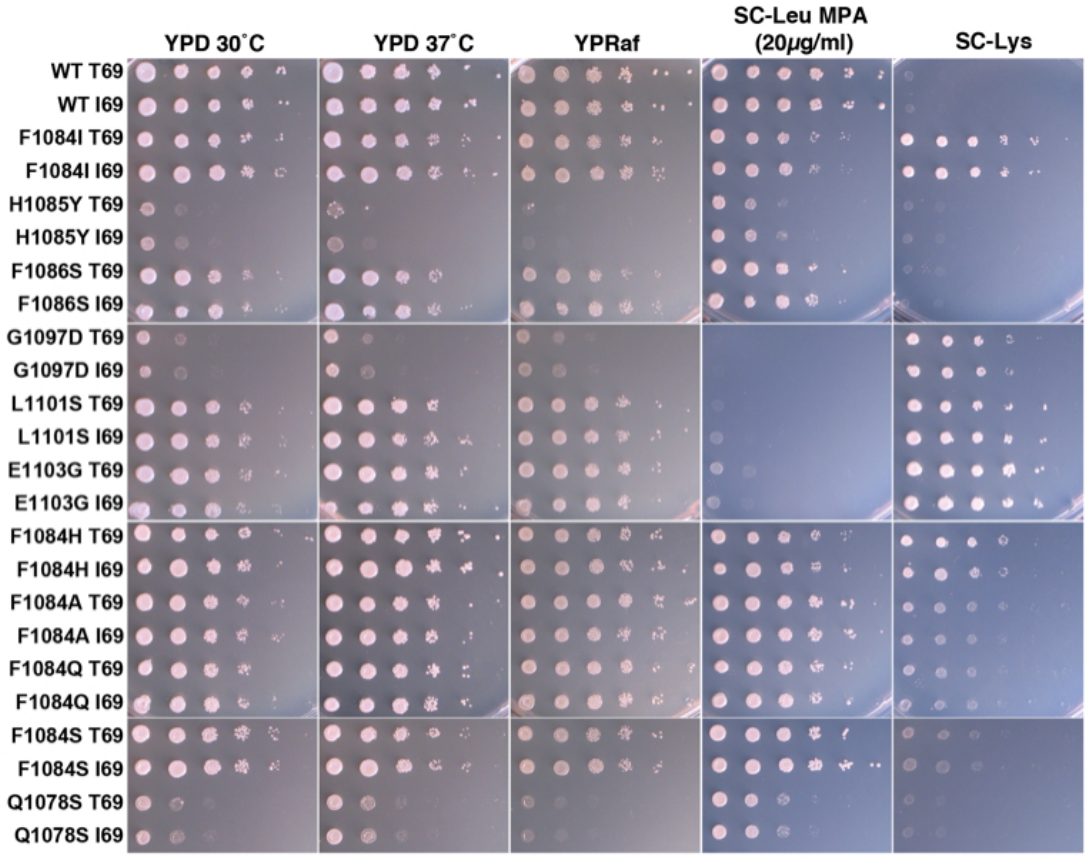


Figure 2-16. No effect of Rpb1 I69 substitution on *in vivo* phenotypes.
10-fold serial dilutions of yeast strains expressing T69 or I69 variants of Rpb1 as the sole source of Rpb1 were spotted onto various media to determine phenotypes. Results indicate that I69 is phenotypically inert.

Yeast Strains, Plasmids And Media

Yeast media are prepared in standard fashion as described in (Amberg et al., 2005) with minor alterations. Yeast extract (1%)(BD), peptone (2%)(BD), dextrose (2%)("YPD") solid medium (2% bacto-agar, BD) is supplemented with adenine (0.15 mM final) and tryptophan (0.4 mM final)(Sigma-Aldrich). YP plates with alternate carbon sources such as raffinose (2%, USB) or raffinose (2%) plus galactose (1%, Sigma-Aldrich) also contain antimycin A (1 µg/ml, Sigma-Aldrich). Minimal media plates are synthetic complete ("SC"/ "Hopkins mix") with amino-acids dropped out as appropriate as described in (Amberg et al., 2005) with minor alterations: per standard batch formulation Adenine Hemisulfate was 2g, Uracil was 2g, myo-inositol was 0.1g, p-Aminobenzoic acid (PABA) was 0.2g. For studies with mycophenolic acid (MPA, Sigma-Aldrich), MPA was added to minimal SC-Leucine medium at 20 µg/ml from a 10 mg/ml stock in ethanol.

Pol II Purification And In Vitro Transcription Assays

Pol II enzymes were purified from yeast strains expressing mutant *rpb1* genes from a low copy plasmid from the endogenous *RPO21* promoter via a tandem-affinity tag (TAP) on Rpb3, in a procedure derived from (Puig et al., 2001) and as described in (Kaplan et al., 2008). All mutant enzymes are Rpb1 I69 variants, except H1085Q/E1103G, F1086S/E1103G and Q1078A/E1103G, which are T69 variants. Transcription reactions of Pol II elongation complexes formed on nucleic acid scaffolds follow the approach described by Komissarova *et al* (Komissarova et al., 2003), and are performed as in (Kaplan et al., 2008) with the following modifications: for some

experiments, amounts of all nucleic acids were reduced to 1/10th the amount stated in (Kaplan et al., 2008) (to 30 pmol from 300 pmol, *etc*.)(Text S1). Reactions were separated on 13.5% acrylamide-bisacrylamide gels (19:1 ratio) containing 1× TBE and 7 M urea and quantitated as previously described (Kaplan et al., 2008).

Primer Extension And Northern Blotting Analysis

Total yeast RNA was purified as described (Schmitt et al., 1990). Primer extension analysis was performed exactly as described (http://labs.fhcrc.org/hahn/Methods/mol_bio_meth/primer_ext.html) (Ranish and Hahn, 1991) with the following modifications. Total RNA used was 30 μg, and in the case of more dilute RNA samples, volumes of reactions were increased by 50% to accommodate greater volume of sample. Reverse transcriptase (RT) was M-MLV RT from either Life Technologies or Fermentas. RT synthesis reactions were supplemented with 1 μl RNAse Inhibitor (Fermentas). Extension products were separated on either 7% or 10% acrylamide-bisacrylamide gels (19:1 ratio) containing 1× TBE and 7 M Urea. Northern blotting was performed essentially as described in manual for GeneScreen hybridization membranes (Perkin-Elmer) with the following modifications. RNA samples (20 μg) were prepared in NorthernMax loading buffer (Ambion/AB). Prehybridization solution did not contain SSPE or SSC buffers, but contained 5× Denhardt's solution, 50 mM Tris-HCl pH 7.5, 1 M NaCl, 0.1% sodium pyrophosphate, 0.1% SDS instead of 1%, 10% Dextran Sulfate, 50% formamide, and 500 μg/ml sheared/denatured salmon sperm DNA. Probes for northern blots were radiolabeled using α-^{32}P-dATP by random priming using the Decaprime II kit (Ambion) according to manufacturer's instructions. Blots were

washed at twice for 10 minutes each wash at 42˚C with 2× SSC, 0.5% SDS, then twice at 67˚C with 5× SSC, 0.5% SDS for 30 minutes each wash, then twice in 0.2× SSC for 30 minutes each wash at room temperature. Primer extension gels and northern blots were visualized by phosphorimaging (GE Healthcare or Bio-Rad) and quantified using ImageQuant 5.0 (GE) or Quantity One (Bio-Rad) software, with data exported to Microsoft Excel for management. Oligo sequences for site-directed mutagenesis, primer extension analysis, amplification of DNA for Northern blotting and *in vitro* transcription are available upon request.

From structure to systems: high-resolution, quantitative genetic analysis of RNA Polymerase II

Summary

RNA polymerase II (RNAPII) lies at the core of dynamic control of gene expression. Using 53 RNAPII point mutants, we generated a point mutant epistatic miniarray profile (pE-MAP) comprising ~60,000 quantitative genetic interactions in Saccharomyces cerevisiae. This analysis enabled functional assignment of RNAPII subdomains and uncovered connections between individual regions and other protein complexes. Using splicing microarrays and mutants that alter elongation rates in vitro, we found an inverse relationship between RNAPII speed and in vivo splicing efficiency. Furthermore, the pE-MAP classified fast and slow mutants that favor upstream and downstream start site selection, respectively. The striking coordination of polymerization rate with transcription initiation and splicing suggests that transcription rate is tuned to

regulate multiple gene expression steps. The pE-MAP approach provides a powerful strategy to understand other multifunctional machines at amino acid resolution.

Introduction

Alterations within a genome often cause specific as well as global phenotypic changes to a cell. Combining two alterations in the same cell allows for measurement of the genetic interaction between them: negative genetic interactions (synthetic sick/lethal) arise when two mutations in combination cause a stronger growth defect than expected from the single mutations. This is often observed for factors participating in redundant pathways or as non-essential components of the same essential complex. In contrast, positive interactions occur when the double mutant is either no sicker (epistatic) or healthier (suppressive) than the sickest single mutant (Beltrao et al., 2010) and may indicate that the factors are components of a non-essential complex and/or perform antagonizing roles in the cell. However, single genetic interactions are often difficult to interpret in isolation; an interaction pattern for a given mutation can be more informative as it reports on the phenotype in a large number of mutant backgrounds (Schuldiner et al., 2005; Tong et al., 2004). These genetic profiles provide highly specific readouts that can be used to identify genes that are functionally related (Beltrao et al., 2010).

One of the first organisms to be genetically interrogated on a large scale was *S. cerevisiae*, where non-quantitative genetic interaction data could be collected using the SGA (synthetic genetic array) (Tong et al., 2004) or dSLAM (heterozygous diploid-based synthetic lethality analysis on microarrays) (Pan et al., 2004) approaches. We developed a technique, termed E-MAP (epistatic miniarray profile) (Collins et al., 2010;

Schuldiner et al., 2005; Schuldiner et al., 2006), which utilizes the SGA methodology and allows for the quantitative collection of genetic interaction data on functionally related subsets of genes, including those involved in chromatin regulation (Collins et al., 2007b), RNA processing (Wilmes et al., 2008), signaling (Fiedler et al., 2009) or plasma membrane function (Aguilar et al., 2010). However, the vast majority of systematic genetic screening interrogates deletions of non-essential genes or hypomorphic knockdown alleles of essential genes. Since many genes, especially essential ones, are multi-functional, these methods perturb all activities associated with a given gene product.

Here, we describe an important advance of the E-MAP approach, which allows us to address higher levels of complexity by examining the genetic interaction space of point mutant alleles of multi-functional genes in a technique we term point mutant E-MAP (or pE-MAP). This method greatly increases the resolution achievable by gene function analysis, as it allows assignment of specific genetic relationships to individual residues and domains. We initially applied this system to *S. cerevisiae* to genetically dissect RNA polymerase II (RNAPII), due to its central role in the control of gene expression. It is an attractive but challenging target, as it is subject to multiple layers of regulation, and the composition of RNAPII transcribing complexes themselves are dynamic throughout the transcription cycle. In addition to querying transcription *per se*, we were hopeful that this approach would identify factors that regulate crosstalk between transcription and co-transcriptional processes. Finally, we could apply

knowledge of this enzyme from *in vitro* biochemical analysis to aid our interpretation of *in vivo* genetic interaction profiles.

In this study, we have used a pE-MAP to functionally dissect RNAPII using alteration-of-function alleles in 5 different subunits of the enzyme. We assign transcriptional activity and specific functions to different residues and regions of RNAPII. By examining the relationship between transcription rate and genetic interaction partners, transcription rate-sensitive factors were revealed. Through the characterization of multiple stages of gene regulation, including start site selection, transcription elongation rate and mRNA splicing, our pE-MAP technique has provided both global and specific insight into structure-function relationships of RNAPII. We propose this strategy as a useful paradigm for the high-resolution interrogation of any multi-functional protein.

Results

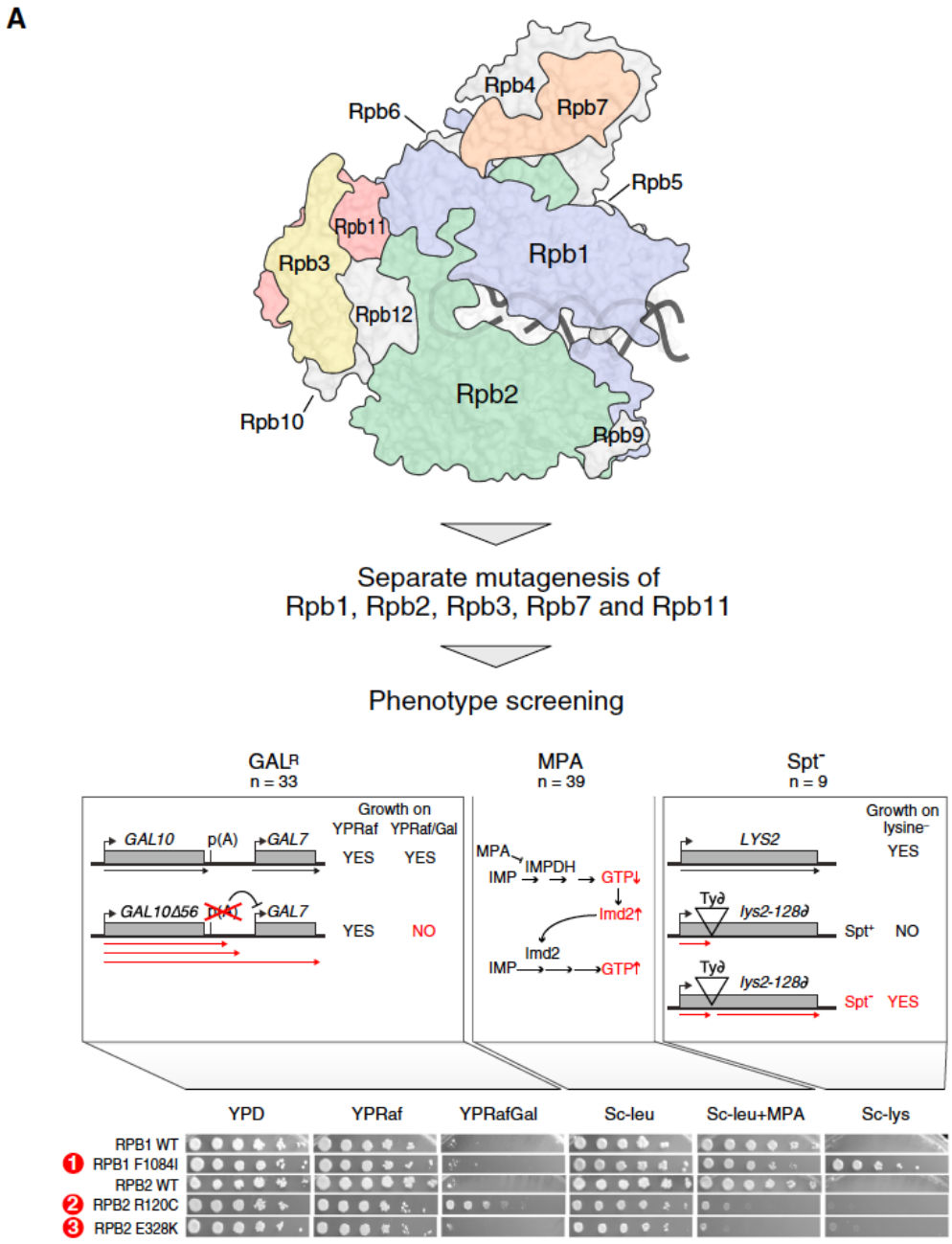
A Set Of Alleles For The Functional Dissection Of RNAPII

To identify residues important for transcriptional regulation *in vivo*, we isolated RNAPII alleles that confer one or more of the following transcription-related phenotypes: suppression of galactose sensitivity in *gal10Δ56* (Gal^R) (Greger and Proudfoot, 1998; Kaplan et al., 2005), the Spt^- phenotype (Winston and Sudarsanam, 1998) or mycophenolic acid (MPA) sensitivity (Shaw and Reines, 2000) (more detail in the legend of **Figure 2-17**, Methods). Each of these phenotypes relates to a gene-specific transcription defect that can be monitored using plate assays (**Figure 2-17** A). Random

mutagenesis by PCR was carried out on the entire coding regions of RNAPII subunit genes *RPB2*, *RPB3*, *RPB7* and *RPB11* and most of *RPO21/RPB1* (Methods). These genes encode the essential subunits unique to RNAPII (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 are shared with RNAPI and RNAPIII, and Rpb4 and Rpb9 are non-essential) (Archambault and Friesen, 1993). In total, 53 single point mutants were identified that exhibit at least one of these phenotypes (Kaplan et al., 2012) (**Figure 2-17** B, **Figure 2-18**).

Analysis of the distribution of phenotypes relative to the RNAPII structure suggested that our alleles might be diverse in their functions. Gal^R and MPA sensitive mutations were broadly distributed, while those with the Spt^- phenotype were less common and more localized (**Figure 2-17** B). Our screens identified mutations in highly conserved residues and structural domains known to be important for RNAPII activity, including the Rpb1 trigger loop, the Rpb1 bridge helix and the Rpb2 lobe and protrusion (Cramer et al., 2001; Gnatt et al., 2001; Kaplan, 2013).We hypothesized that quantitatively measuring the genetic interactions of specific residues might provide insight into the functions of these RNAPII regions, and might be useful for identifying protein-protein interaction interfaces.

Figure 2-17. Generation and selection of RNAPII point mutants.
A. RNAPII point mutants were screened for three transcription-related phenotypes. Gal^R, left: Deletion of the major *GAL10* pA site (*gal10*D*56*) results in RNAPII read-through and interference with initiation at *GAL7*, causing a Gal-sensitive phenotype (Greger and Proudfoot, 1998). Gal^R mutants increase *GAL10* 3'-end formation/termination, thereby rescuing *GAL7* expression. MPA, middle: Mycophenolic acid (MPA) inhibits IMP-dehydrogenase (IMPDH)-dependent GTP synthesis, but is counteracted by upregulation of an MPA-resistant form of IMPDH, *IMD2*. Transcriptional defects that are sensitive to low GTP levels or reduce *IMD2* expression render cells sensitive to MPA (Shaw and Reines, 2000). Spt^-, right: Insertion of a Ty-retrotransposon into *LYS2* (*lys2-128*d*)* results in a lysine auxotrophy due to transcription block. Certain mutants suppress *lys2-128*d and allow expression of *LYS2*, due to activation of an internal promoter (Winston and Sudarsanam, 1998). Spot tests to identify each phenotype for three representative mutants are displayed. **B**. Positions, mutations and phenotypes of the 53 single point mutants analyzed in the pE-MAP (**Figure 2-18**). Colored lines represent subunit sequences, with mutations denoted by residue numbers and single letter amino acid codes for WT and mutant. See also **Figure 2-18**.


A
Rpb4
Rpb7
Rpb6
Rpb5
Rpb11
Rpb1
Rpb3
Rpb12
Rpb2
Rpb10
Rpb9
Separate mutagenesis of
Rpb1, Rpb2, Rpb3, Rpb7 and Rpb11
Phenotype screening
GALR
n = 33
MPA
n = 39
Spt-
n = 9
Growth on
YPRaf YPRaf/Gal
GAL10 p(A) GAL7
YES YES
GAL10Δ56 GAL7
YES NO
MPA
IMP IMPDH GTP↓
Imd2↑
Imd2
IMP GTP↑
Growth on
lysine-
LYS2
YES
Tyδ lys2-128δ
Spt+ NO
Tyδ lys2-128δ
Spt- YES
YPD
YPRaf
YPRafGal
Sc-leu
Sc-leu+MPA
Sc-lys
RPB1 WT
1 RPB1 F1084I
RPB2 WT
2 RPB2 R120C
3 RPB2 E328K

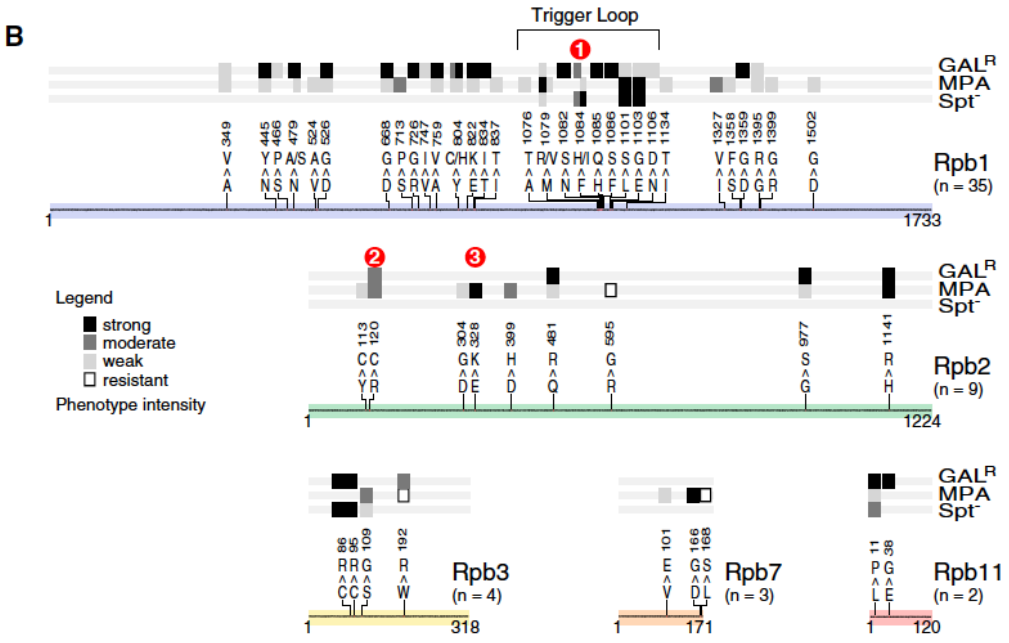


Figure 2-17 Continued**.**

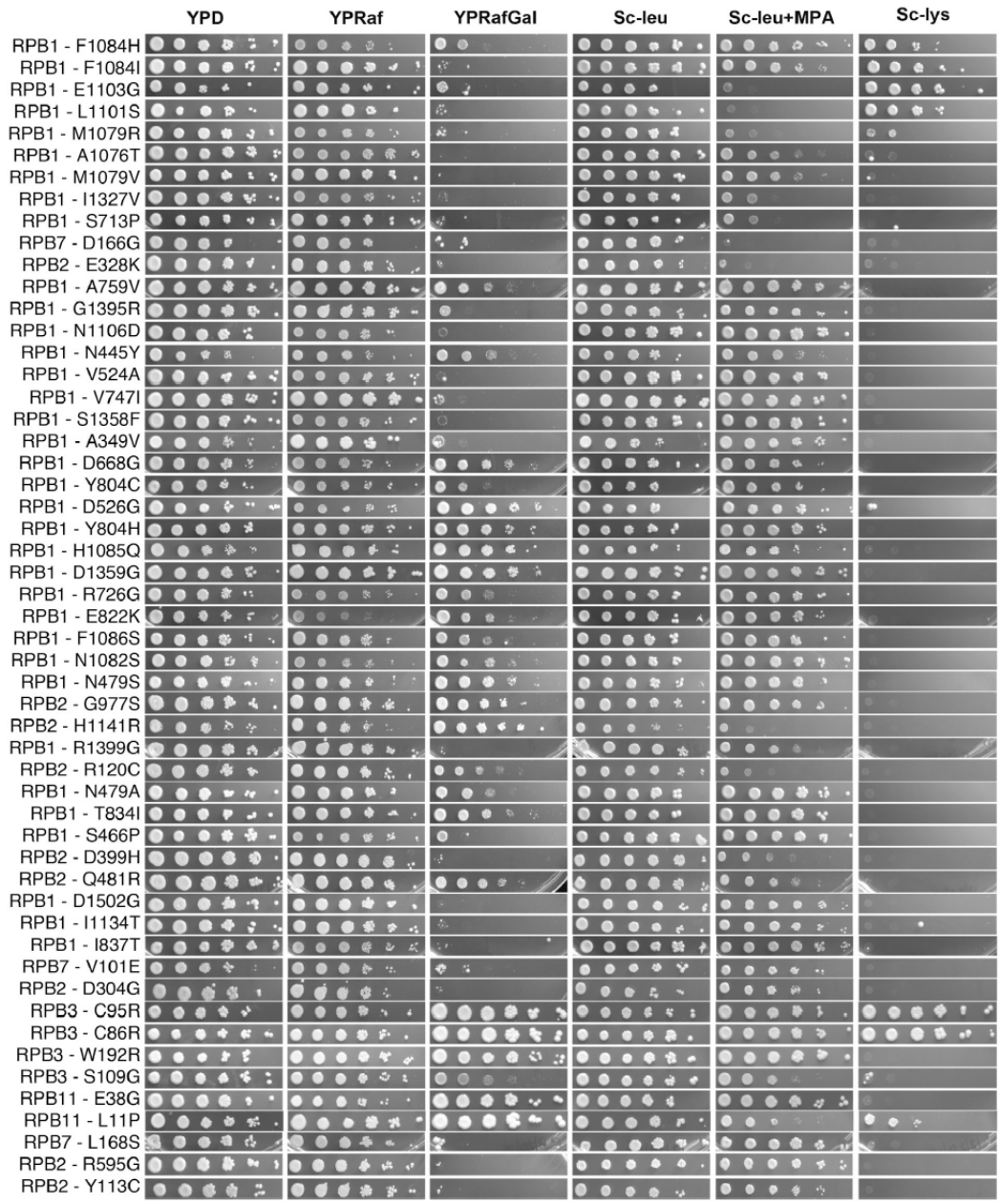


Figure 2-18. Complete collection of spot tests for identification of Gal^R, MPA sensitivity and Spt$^-$ phenotypes.

Spot tests on relevant media were carried out to identify RNAPII mutants that exhibit at least one of the three phenotypes, as described in **Figure 2-17** and **Methods**. The order of RNAPII mutants follows that of the clustered pE-MAP (**Figure 2-19**).

An RNAPII Point Mutant Epistatic Miniarray Profile (pE-Map)

The 53 RNAPII point mutants (**Figure 2-17 B**) were crossed against ~1200 deletion and DAmP (Decreased Abundance by mRNA Perturbation) alleles (Schuldiner et al., 2005), which represent all major biological processes. We thus created a quantitative pE-MAP comprising approximately 60,000 double mutants (http://interactome-cmp.ucsf.edu). Two-dimensional hierarchical clustering of these data effectively grouped together genes from known complexes and pathways based on their interactions with the point mutants (**Figure 2-19**). Our previous studies have demonstrated that genes encoding proteins that physically associate often have similar genetic interaction profiles (Collins et al., 2007b; Roguev et al., 2008). We wondered whether similar observations would characterize our point mutant data, even though we have only examined 5 subunits of a single molecular machine. To test this, we generated a receiver operating characteristic (ROC) curve that measures how well the genetic profiles of the deletion and DAmP mutants in our pE-MAP predict known physical interactions between their encoded proteins (**Methods**). We found that the predictive power of the pE-MAP was similar to that of a previously published E-MAP (Collins et al., 2007b), indicating that the genetic interactions of the RNAPII point mutants report on connections among virtually all cellular processes (**Figure 2-20** A, **Figure 2-19**).

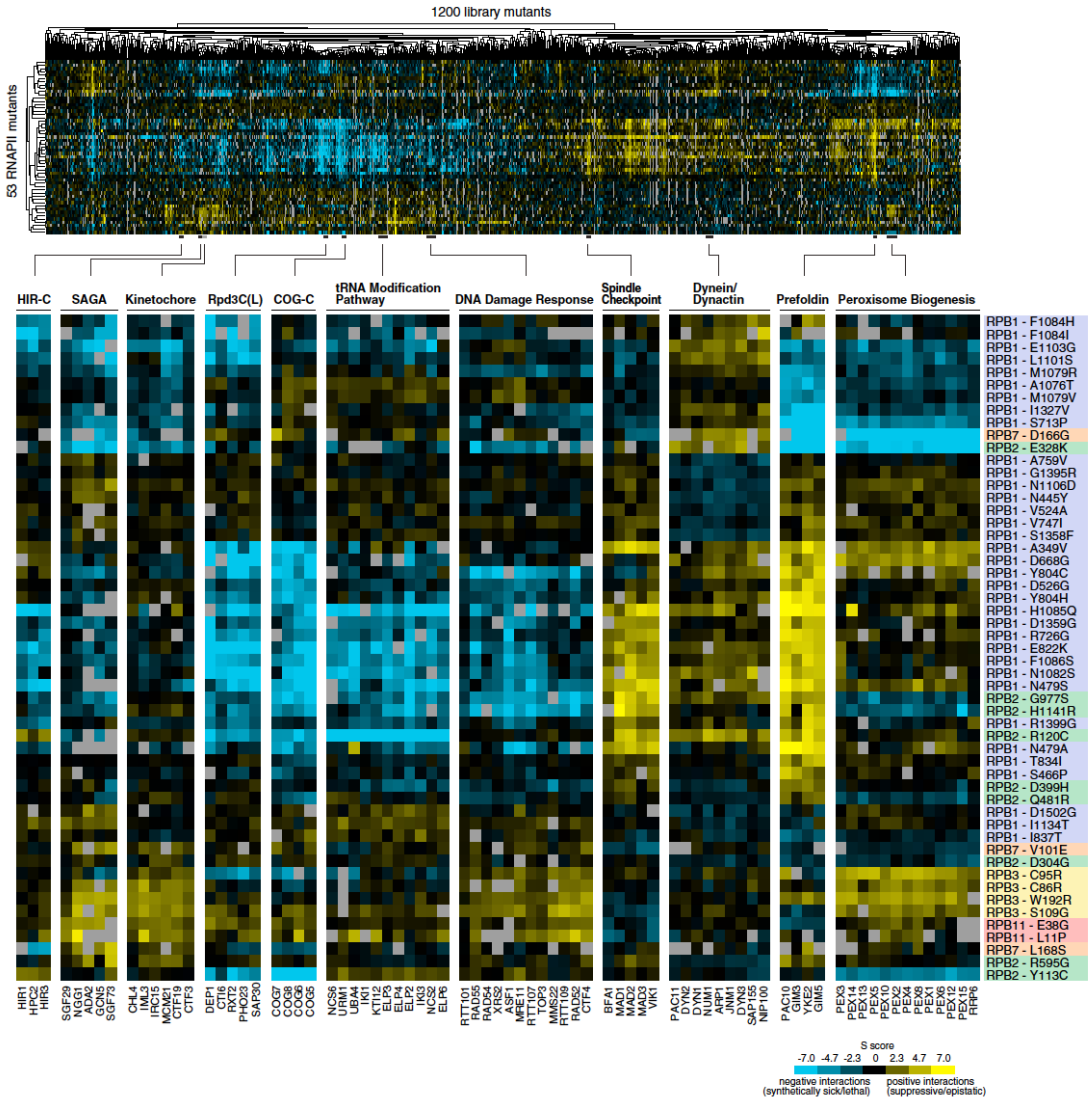


Figure 2-19. A high-resolution genetic interaction map of RNAPII point mutants. The 53 point mutants described in **Figure 2-17** B were genetically screened against a library of 1200 mutants carrying single gene deletions or DAmP alleles. The resulting pE-MAP reports on a total of 59,534 genetic interactions between single residue point mutations and deletions of non-essential genes or DAmP alleles of essential genes. The dendrograms organize the mutants functionally and were generated by 2-D hierarchical clustering. Several representative clusters of library mutants that belong to the same complex or pathway are highlighted. The background colors of RNAPII mutant labels correspond to subunit and match the color scheme in **Figure 2-17**.

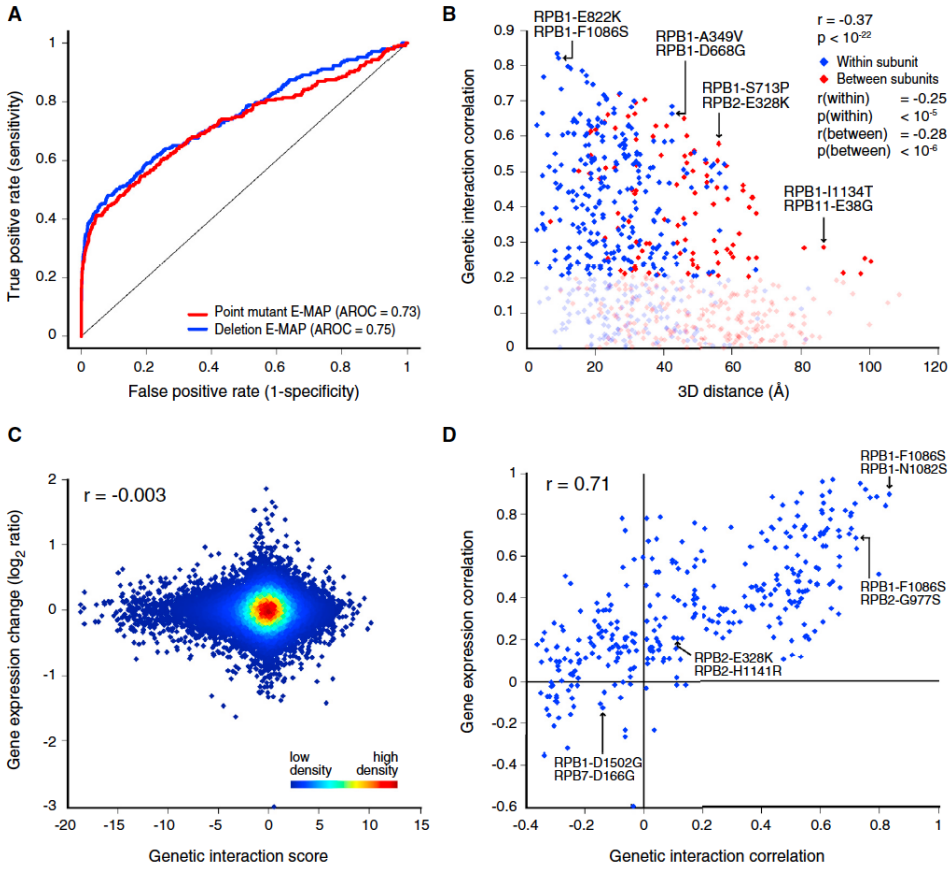


Figure 2-20. pE-MAP interactions span numerous biological processes.
pE-MAP interactions depend on spatial location of mutated residues and are not direct consequences of changes in gene expression. **A**. ROC curves comparing the power of genetic profile correlations from the pE-MAP (red) and an E-MAP focused on chromosome biology (Collins et al., 2007b) (blue) to predict physical interactions between pairs of proteins (Collins et al., 2007a) (**Methods**). **B**. Genetic profile correlations between pairs of mutated residues compared to the three-dimensional distance between their α-carbons (**Table S3**). Blue points denote residue pairs within the same RNAPII subunit; red points represent pairs in different subunits. Negatively correlated residue pairs were excluded, as were four mutants of residues absent from the coordinate file (PDB ID 2E2H) (Rpb1 D1502, Rpb7 V101, Rpb7 D166, Rpb7 L168) (Wang et al., 2006). Correlations between 0 and 0.2 are dimmed to highlight trends at higher correlations. **C**. Effect of RNAPII point mutations on gene expression, compared to the corresponding genetic interaction scores between RNAPII mutants and deletion/DAmP alleles. All combinations of the 26 RNAPII mutants and 1192 library genes/mutants that were examined via both pE-MAP and expression analyses are included. No global changes in gene expression were observed (measured by spike-in control RNA, **Methods**). **D**. Comparison between pairwise RNAPII mutant correlations of genetic interaction profiles and gene expression profiles.

Next, to examine whether the spatial location of a mutated residue is a determinant of its function, we compared the similarity of pairs of RNAPII genetic profiles to the three-dimensional distance between the mutated residues (Wang et al., 2006). We observed a strong correlation ($r=-0.37$, $p<10^{-22}$) and the trend is significant both for residue pairs residing in the same subunit ($r=-0.25$, $p<10^{-5}$) and for those in different subunits ($r=-0.28$, $p<10^{-6}$) (**Figure 2-20** B). This suggests that structural proximity correlates with functional similarity and that high resolution genetic interaction profiling could provide information for targets whose structures have not yet been determined.

Comparison Of Genetic And Gene Expression Profiles Derived From The RNAPII Alleles

To determine if any given genetic interaction might result from the point mutation affecting the expression of the corresponding gene, we subjected 26 of the RNAPII mutants to genome-wide gene expression analysis (GEO accession number: GSE47429). We found no correlation ($r=-0.003$) between an RNAPII mutant's genetic interaction score with a gene deletion or DAmP allele and the expression change of that gene due to the RNAPII mutation (**Figure 2-20** C). Therefore, connections must be due to more complex relationships between the mutated residues and the library genes. Nonetheless, these datasets allowed us to test whether the clustering of the RNAPII mutants in the pE-MAP (**Figure 2-19**) could be recapitulated using their gene expression profiles. We thus assessed pair-wise RNAPII mutant similarity based on genetic and gene expression profiles separately and found that these two measures are highly correlated ($r=0.71$, **Figure 2-20** D). Therefore, these orthogonal datasets provide a common biological

framework for functionally organizing the RNAPII mutants, allowing us to study the underlying biology behind these mutants and their phenotypes.

Functional Associations Between RNAPII Residues And Protein Complexes

In an effort to link individual RNAPII mutations to specific cellular functions, data from the pE-MAP was compared to a published genetic interaction dataset containing profiles from >4000 genes (Costanzo et al., 2010). These genes were classified based on complex membership of their encoded proteins (Benschop et al., 2010) and Mann-Whitney U statistics were used to identify RNAPII mutants with high profile similarity to members of specific complexes (**Methods**, **Figure 2-21** A). We uncovered a number of connections, including several point mutants having similar genetic profiles to mutants of components of Mediator and the Rpd3C(L) histone deacetylase complex (Carrozza et al., 2005; Keogh et al., 2005). Unexpectedly, we also observed that several RNAPII mutants are significantly correlated genetically to kinetochore mutants (**Figure 2-21** A). We carried out chromosome transmission fidelity (CTF) assays on 19 of our RNAPII mutants, including those linked specifically to the kinetochore (**Methods**) (Spencer et al., 1990). Only four of the tested mutants exhibited chromosome loss in more than 15% of their colonies and these were genetically linked to the kinetochore in our analysis and had similar genetic profiles to each other in our pE-MAP (**Figure 2-21** A, B). Recent studies indicate that a certain level of transcription by RNAPII at the centromere is required for centromere function and high-fidelity chromosome segregation in budding yeast (Ohkuni and Kitagawa, 2011). Using specific constructs designed to ascertain centromere sensitivity to transcriptional read-through, we did not

observe any defects in kinetochore integrity in these RNAPII mutants (data not shown). Ultimately, further work will be required to understand the connection between these RNAPII point mutants and chromosome segregation. A full point mutant module map from alleles of all subunits is presented in **Figure 2-22**.

PE-Map Identifies Alleles Involved In Start Site Selection And Can Finely Distinguish Between Different Phenotypic Categories

What other transcription defects might underlie differences in the genetic profiles of specific RNAPII alleles? Our recent work has shown that mutations in the Rpb1 trigger loop (TL), a dynamic element in the active site that couples correct NTP substrate recognition with catalysis, can alter transcription start site selection *in vivo*. For example, *rpb1* E1103G shifts transcription start site selection upstream at *ADH1* whereas *rpb1* H1085Y shifts distribution of start sites downstream (Kaplan et al., 2012). The pE-MAP subcluster containing E1103G includes an additional 10 mutants in *RPB1*, *RPB2* and *RPB7* (**Figure 2-23** A). We examined 8 of these for defects in start site selection at *ADH1* by primer extension and found that, like E1103G, all had a preference for upstream start site selection (**Figure 2-23** B, **Figure 2-24** A), consistent with their clustering with E1103G (**Figure 2-23** A). Four of these mutants are also in the TL (*rpb1* F1084I, M1079R, A1076T, and M1079V); however, two were in other regions of Rpb1 (*rpb1* I1327V and S713P). Further inspection reveals these are in close proximity to the TL (**Figure 2-23** C) (Kettenberger et al., 2004; Wang et al., 2006), suggesting they too may directly regulate the active site. Interestingly, the other 2 mutants tested are not close to the TL (*rpb2* E328K and *rpb7* D166G); these mutations may allosterically

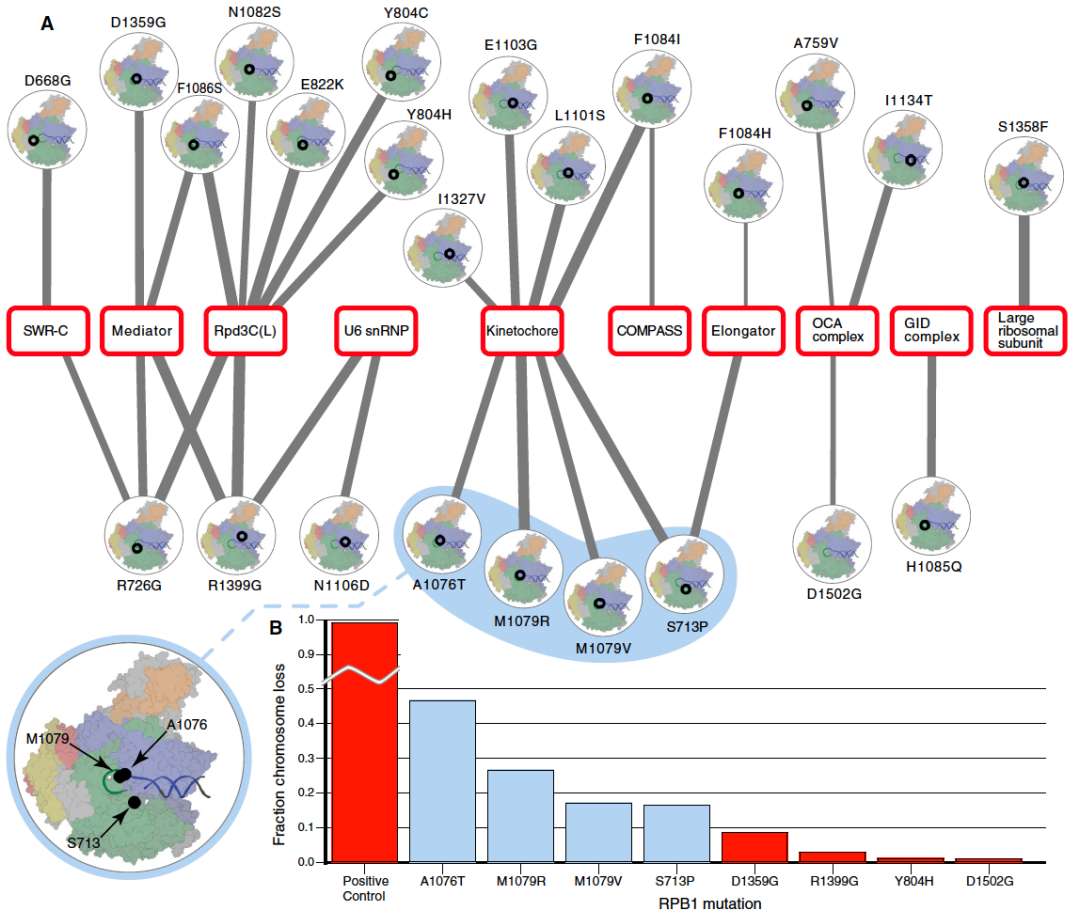


Figure 2-21. Comparison of the pE-MAP with previous collected genetic interaction data reveals functional associations between RNAPII residues and protein complexes.

A. Module map depicting genetic similarity of RNAPII mutants with genes encoding the indicated protein complex subunits (**Methods**). Edge widths correspond to statistical significance of connections. Only *RPB1* edges with a false discovery rate <0.1 are displayed. Four mutated residues linked to the kinetochore are highlighted in blue and the blowup indicates their structural locations. **B**. 19 mutants were examined using a chromosome transmission fidelity (CTF) assay. The four kinetochore-linked mutants highlighted in (**A**) exhibit chromosome loss in more than 15% of their colonies (blue bars), whereas unlinked mutants display no or weak phenotype (red bars, representative set).

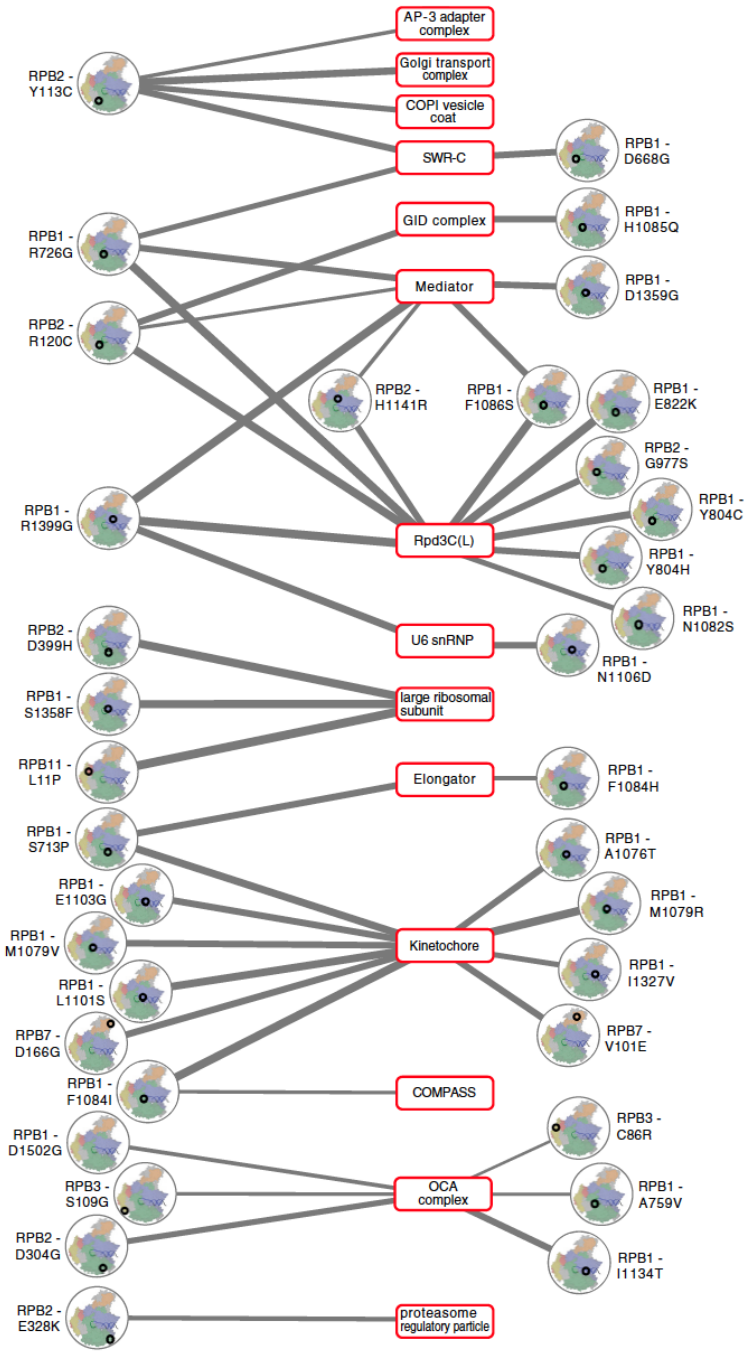
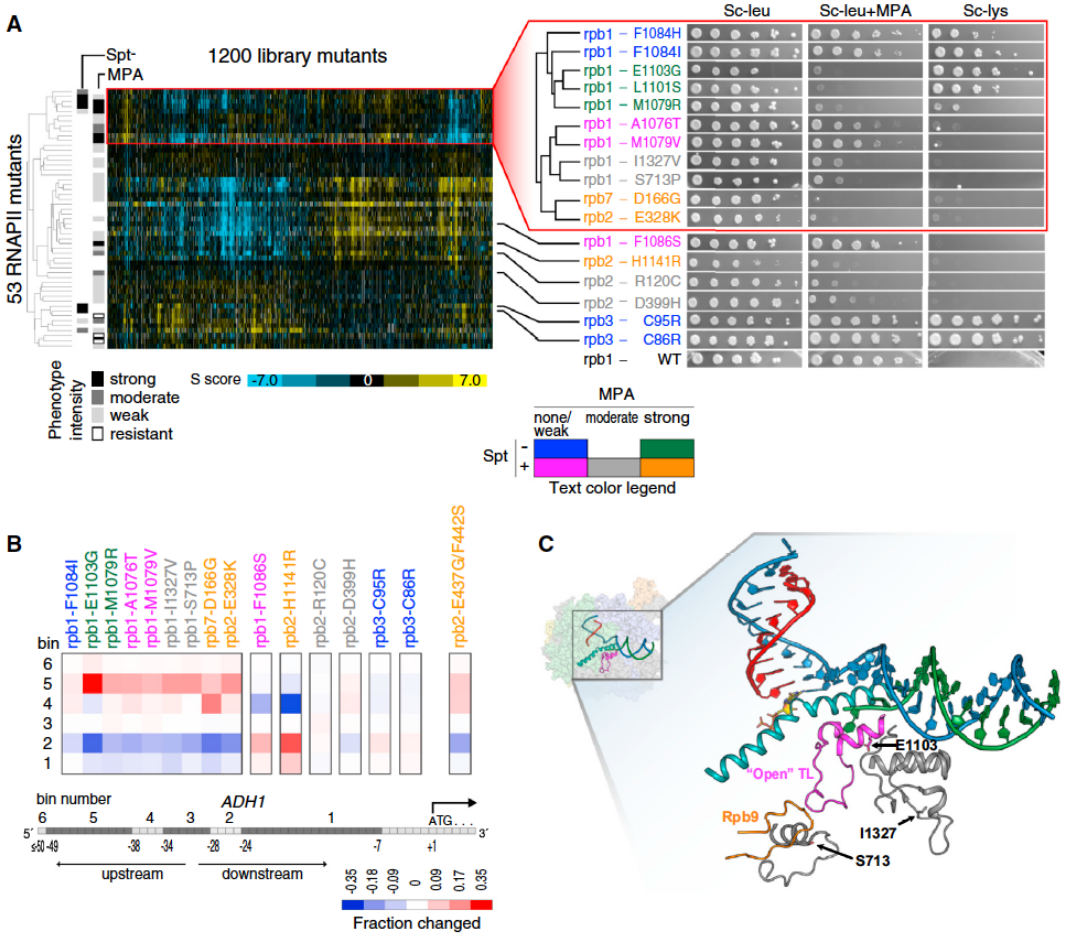


Figure 2-22. Functional connections between RNAPII mutants and protein complexes.

Functional associations between RNAPII mutants and protein complexes were determined as described in **Figure 2-21** and **Methods**. Edge widths correspond to the statistical significance of connections, and only edges with a false discovery rate (FDR) <0.1 are displayed.

Figure 2-23. pE-MAP profiles differentiate between subtle changes in transcription-related phenotypes and identify RNAPII mutations that affect start site selection.

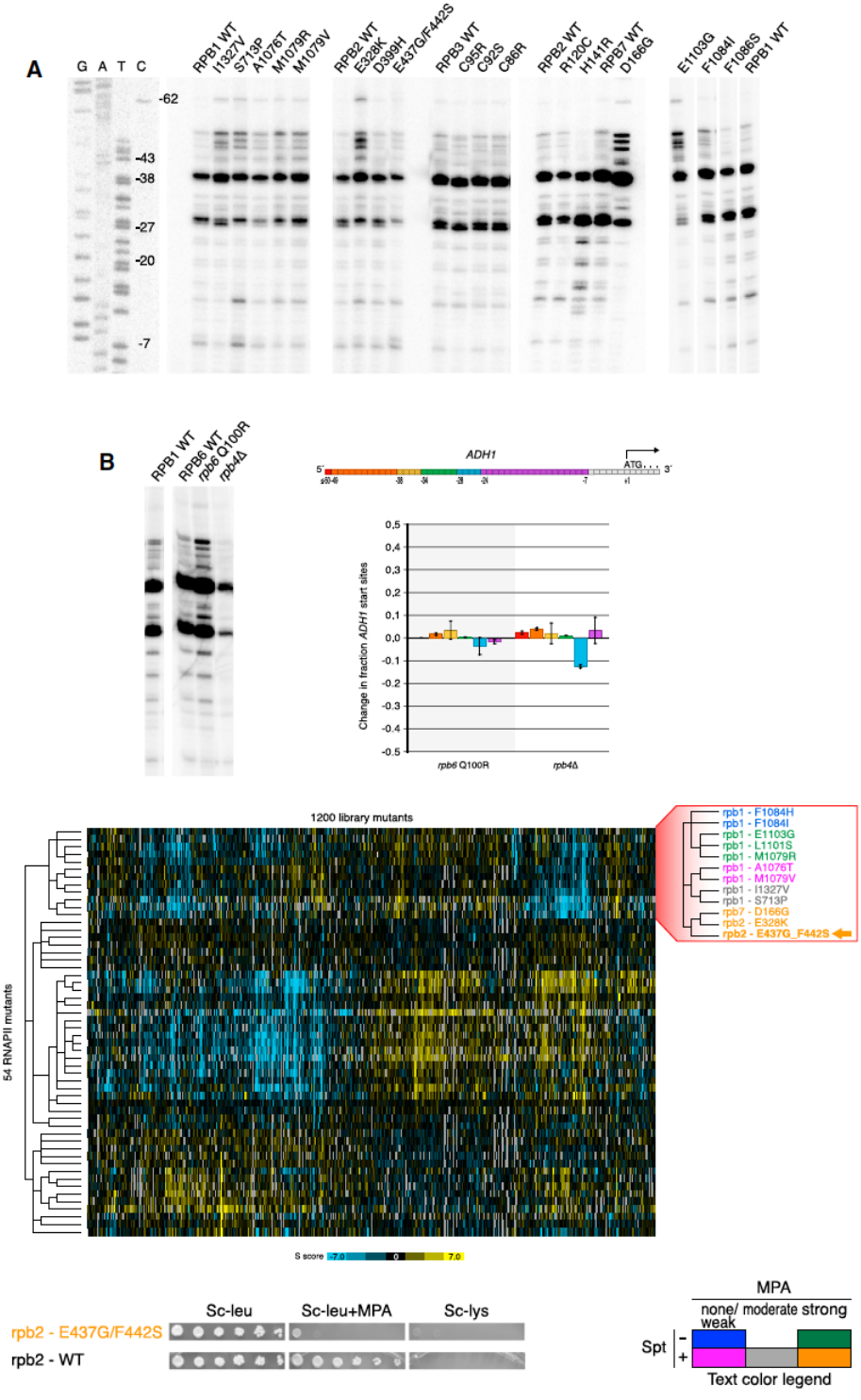
A. pE-MAP clustering in relation to MPA and Spt- phenotypes of alleles. The RNAPII alleles are clustered by pE-MAP profiles and their colors indicate degrees of MPA and Spt- phenotypes (determined from the spot tests). **B**. Effect of RNAPII mutations on start site selection at *ADH1* determined by primer extension analysis. The heatmap describes the fractional change of start site in each bin of the *ADH1* schematic (bottom). **C**. Rpb1 I1327 and Rpb1 S713 connect to the TL (magenta). Mutations in I1327 could affect the structural region of the TL (E1103) via a network of loops and helices in Rpb1 (grey), while S713 is close to the TL catalytic site in its open conformation, via a Rpb9 loop (orange). In particular, the proline substitution, S713P, could result in structural changes affecting the TL. Coordinates for TL, Rpb9 and the S713 loop are from PDB ID 1Y1V (Kettenberger et al., 2004) and all other from 2E2H (Wang et al., 2006). The bridge helix is shown in cyan, template DNA in blue, non-template DNA in green, RNA in red, and the incoming GTP base is colored by atom. See also **Figure 2-24**.Figure 0–1-2-23 pE-MAP profiles differentiate between subtle changes in transcription-related phenotypes and identify RNAPII mutations that affect start site selection.


A
Spt-
MPA
1200 library mutants
53 RNAPII mutants
Sc-leu
Sc-leu+MPA
Sc-lys
rpb1 – F1084H
rpb1 – F1084I
rpb1 – E1103G
rpb1 – L1101S
rpb1 – M1079R
rpb1 – A1076T
rpb1 – M1079V
rpb1 – I1327V
rpb1 – S713P
rpb7 – D166G
rpb2 – E328K
rpb1 – F1086S
rpb2 – H1141R
rpb2 – R120C
rpb2 – D399H
rpb3 – C95R
rpb3 – C86R
rpb1 – WT
Phenotype intensity
strong
moderate
weak
resistant
S score
-7.0
0
7.0
MPA
none/ weak
moderate
strong
Spt
-
+
Text color legend
B
bin
rpb1-F1084I
rpb1-E1103G
rpb1-M1079R
rpb1-A1076T
rpb1-M1079V
rpb1-I1327V
rpb1-S713P
rpb7-D166G
rpb2-E328K
rpb1-F1086S
rpb2-H1141R
rpb2-R120C
rpb2-D399H
rpb3-C95R
rpb3-C86R
rpb2-E437G/F442S
bin number
ADH1
ATG...
upstream
downstream
Fraction changed
C
E1103
"Open" TL
Rpb9
I1327
S713

impact the active site or function independent of the TL by recruiting other factors to the transcription apparatus. Importantly, *rpb7* D166G is the first identified *RPB7* mutation with a start site defect. We additionally examined start site selection in *rpb4*Δ and *rpb6* Q100R, as both are expected to reduce the association of Rpb4/Rpb7 with RNAPII (Edwards et al., 1991; Tan et al., 2003). Neither altered start site selection at *ADH1* (**Figure 2-24** B)**,** indicating the *rpb7* D166G mutant exerts a unique effect on RNAPII function (see **Discussion**). These data provide an example of how mechanistic information on structure-function relationships can be extracted from the pE-MAP.

In our screening process, we had also identified an *rpb2* allele mutated at two residues in close proximity (E437G/F442S) within the tip of the Rpb2 protrusion domain, whose genetic profile also clusters with the upstream start site mutants (**Figure 2-24** C). The Rpb2 lobe and protrusion domains physically contact TFIIF (Chen et al., 2007) and consistent with this, we observed that, similar to what has been reported for TFIIF alleles (Eichner et al., 2010; Ghazy et al., 2004) (**Figure 2-25** B), *rpb2* E437G/F442S is

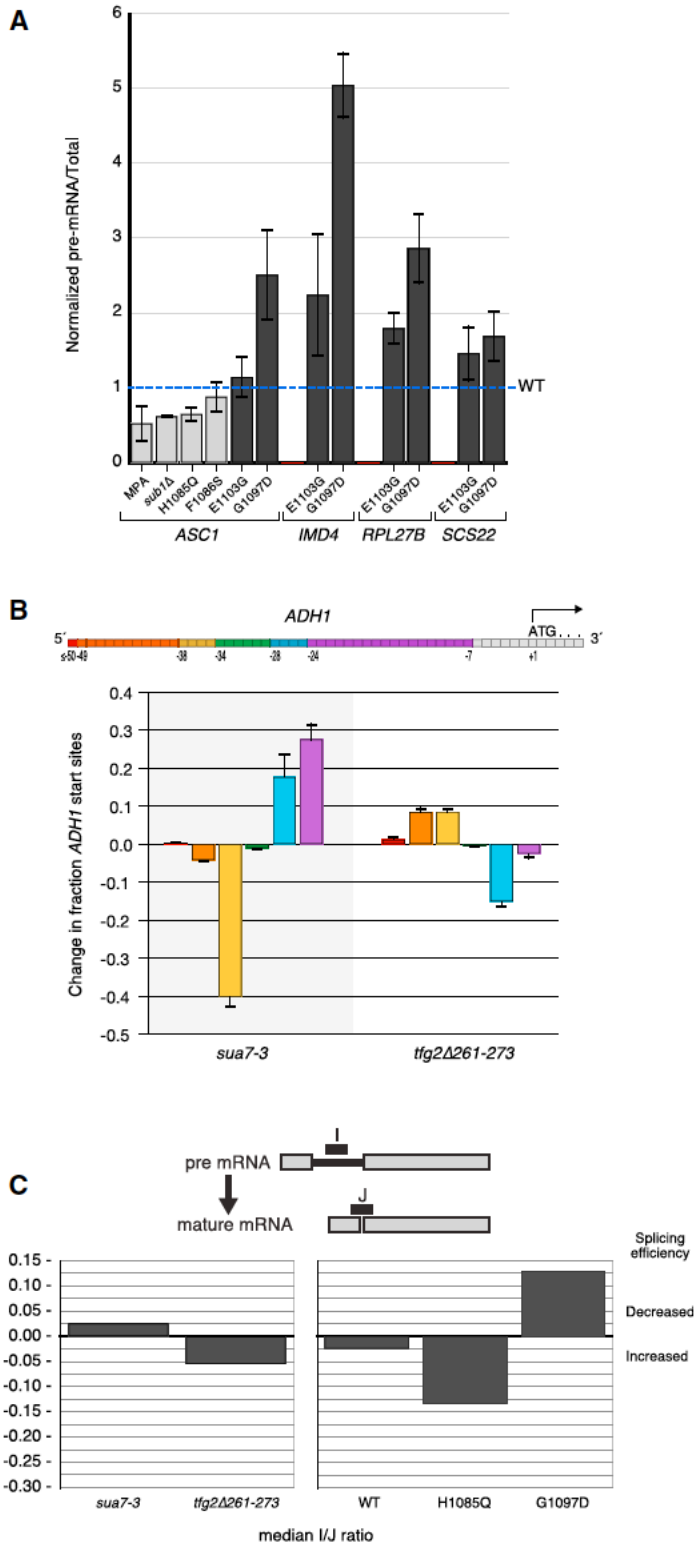
Figure 2-24. Primer extension analysis at *ADH1* to identify the effect of RNAPII mutations on start-site selection, and data relating to *rpb2* E437G/F442S.
A. Sequencing gel for primer extension analysis. Lanes 1-4 correspond to Sanger dideoxy sequencing reactions as reference ladders, and the following lanes carry reverse transcription products from RNAPII alleles as indicated. **B**. Primer extension analysis at *ADH1* to map transcription start sites for *rpb4*Δ and *rpb6* Q100R mutations. The colors of the bars correspond to the sequence windows indicated in the *ADH1* schematic (top) and the heights specify the fraction change of transcription start in the mutant compared to wt. Means and standard deviations were derived from three independent experiments. **C**. Clustering of *rpb2* E437G/F442S in pE-MAP. **D**) MPA sensitivity and Spt phenotypes of *rpb2* E437G/F442S.


A
G A T C
RPB1 WT
I1327V
S713P
A1076T
M1079R
M1079V
RPB2 WT
E328K
D399H
E437G/F442S
RPB3 WT
C95R
C92S
C86R
RPB2 WT
R120C
H141R
RPB7 WT
D166G
E1103G
F1084I
F1086S
RPB1 WT
-62
-43
-38
-27
-20
-7
B
RPB1 WT
RPB6 WT
rpb6 Q100R
rpb4Δ
ADH1
5'
ATG...
3'
Change in fraction ADH1 start sites
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
-0,2
-0,3
-0,4
-0,5
rpb6 Q100R
rpb4Δ
C
1200 library mutants
54 RNAPII mutants
S score
-7.0
0
7.0
rpb1 - F1084H
rpb1 - F1084I
rpb1 - E1103G
rpb1 - L1101S
rpb1 - M1079R
rpb1 - A1076T
rpb1 - M1079V
rpb1 - I1327V
rpb1 - S713P
rpb7 - D166G
rpb2 - E328K
rpb2 - E437G_F442S
D
Sc-leu
Sc-leu+MPA
Sc-lys
rpb2 - E437G/F442S
rpb2 - WT
MPA
none/ moderate strong
weak
Spt
-
+
Text color legend

sensitive to MPA (**Figure 2-24** D) and has a preferential upstream start site selection (**Figure 2-23** B) (see **Discussion**). Additional *rpb2* alleles that confer MPA sensitivity also map near the protrusion (R120C) or to the lobe (D399H), however these did not alter *ADH1* start site selection and are genetically distinct from E1103G or other *rpb2* alleles, illustrating the fine resolving power of the pE-MAP (**Figure 2-23** A, B).

Despite its unbiased nature, the pE-MAP precisely grouped the mutants within the upstream start site cluster based on their Spt^- and MPA phenotypes in the plate assays (**Figure 2-23** A). However, there are other mutants with MPA or Spt^- phenotypes that did not exhibit upstream start preference and, notably, cluster apart from the ones that do. (**Figure 2-23** B, **Figure 2-24** A). These data collectively suggest that the pE-MAP has the resolving power to categorize the mutations causing upstream start site selection and within this group can further separate them into specific Spt^- and MPA phenotypic categories.

Figure 2-25. Confirmation of RNAPII single-mutant splicing phenotypes and interrogation of potential connection between splicing and start site selection. **A.** Pre-mRNA and total mRNA for *IMD4*, *RPL27B* and *SCS22* transcripts were measured in the indicated strains and conditions. Shown are pre-mRNA/total mRNA ratios, normalized to that of an untreated wild-type. Error bars represent standard deviation of 5-6 biological replicates. **B.** Primer extension analysis at ADH1 to map transcription start sites for mutations in TFIIB (*sua7-3*) and TFIIF (*tfg2*Δ*261-273*). The colors of the bars correspond to the sequence windows indicated in the ADH1 schematic (top) and the heights specify the fraction change of transcription start in the mutant compared to WT. Error bars represent SD. Means and SD were derived from three independent experiments. **C.** TFIIB (*sua7-3*) and TFIIF (*tfg2*Δ*261-273*) mutants were analyzed on splicing microarrays. Median I/J log2 ratios for the complete profiles of the TFII mutants are shown in the left panel, with WT, a slow and a fast rpb1 mutant included for reference in the right panel.


A
Normalized pre-mRNA/Total
WT
MPA
sua7Δ
H1085Q
F1086S
E1103G
G1097D
E1103G G1097D
E1103G G1097D
E1103G G1097D
ASC1
IMD4
RPL27B
SCS22
B
ADH1
5′
3′
ATG. . . .
Change in fraction ADH1 start sites
sua7-3
tfg2Δ261-273
C
pre mRNA
mature mRNA
I
J
Splicing efficiency
Decreased
Increased
sua7-3
tfg2Δ261-273
WT
H1085Q
G1097D
median I/J ratio

In Vitro Biochemical Activity Correlates With pE-Map Profiles And Gene Expression

We next focused on the genetic profiles of a series of active-site mutants whose *in vitro* elongation rates range from <0.1 to >2-fold that of WT RNAPII (Kaplan et al., 2012; Kaplan et al., 2008; Malagon et al., 2006). This series allowed us to address questions regarding the *in vivo* consequences of altered elongation rates. Clustering these mutants based on their pE-MAP profiles yielded two distinct subsets that differ by transcription rate: fast (*rpb1* E1103G, L1101S and F1084I) and slow mutants (*rpb1* F1086S, N1082S, N479S and H1085Q) (**Figure 2-26**). The gene expression profiles of these mutants also group them into the same two subsets. Thus, pairs of mutants with similar *in vitro* transcription rates show both highly similar genetic and expression profiles (**Figure 2-26** B), indicating that altered catalytic activity may likely be a defining feature *in vivo.*

We reasoned that because combining two TL mutations that individually increase

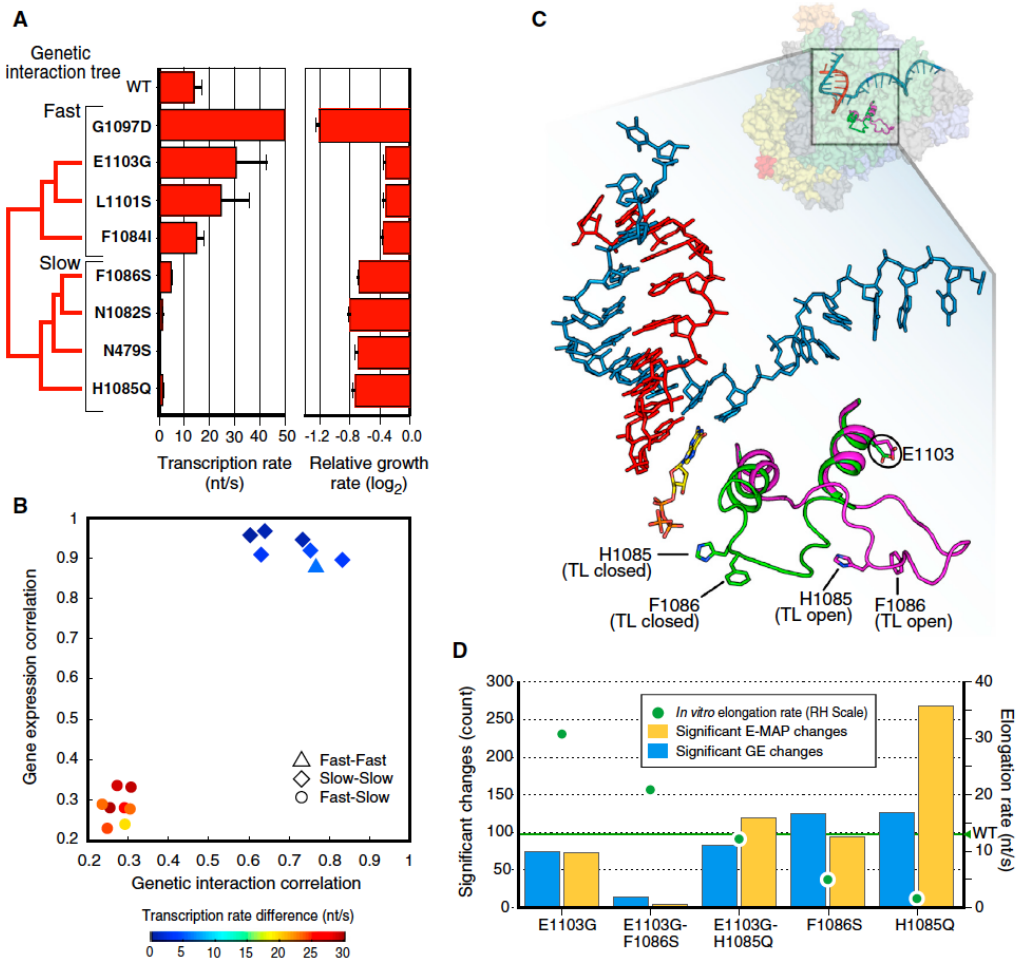


Figure 2-26. pE-MAP and expression profiles are indicative of biochemical activity. **A**. *In vitro* transcription rates from (Kaplan et al., 2012) and *in vivo* growth rates relative to WT for RNAPII active site mutants. The dendrogram was generated via hierarchical clustering of the genetic profiles. Means and standard deviations of growth rates were derived from three technical replicates. Note that *rpb1* G1097D was too sick for reproducible E-MAP analysis. **B**. *In vitro* transcription rate difference between pairs of active site mutants in relation to their genetic and expression profile correlations. **C**. Residues H1085 and F1086 reside in the catalytic site of the TL, while E1103 is part of the distal flanking alpha helix that structurally constrains the TL in open conformations. TL is shown in green (closed) and magenta (open), template DNA in blue, RNA in red and the incoming GTP base is colored by atom. Coordinates for open TL are from PDB ID 1Y1V and all other from 2E2H. **D**. Counts of high-scoring interactions (pE-MAP score >3.3 (97.5 percentile) or <-5.1 (2.5 percentile)) in the complete genetic profiles, or changes >1.7-fold in the genome-wide expression profiles, of the indicated RNAPII mutants. *In vitro* transcription rates are indicated on the right-hand scale.

and decrease elongation rate results in RNAPII with near-WT elongation rate *in vitro* and growth rate *in vivo* (Kaplan et al., 2012) (**Figure 2-27** A), these double mutants would also exhibit near-WT pE-MAP and gene expression profiles. Indeed, when we combine *rpb1* E1103G (fast) and F1086S (slow), we see very few high scoring genetic interactions compared to the single mutants and few genes whose expression changes by more than 1.7-fold (**Figure 2-26** C, D). However, the pE-MAP and gene expression assays had the resolving power to also identify a double mutant (*rpb1* E1103G/H1085Q) that deviated from this general rule (**Figure 2-26** C, D). Despite exhibiting partially suppressive genetic and expression patterns, there are still significant effects that correlate well with E1103G (**Figure 2-27** B), suggesting a more complex genetic relationship with this double mutant. Taken together, our double mutant analyses are consistent with the notion that the genetic interactions and gene expression changes in fast and slow mutants are, for the most part, defined by the catalytic activity of RNAPII.

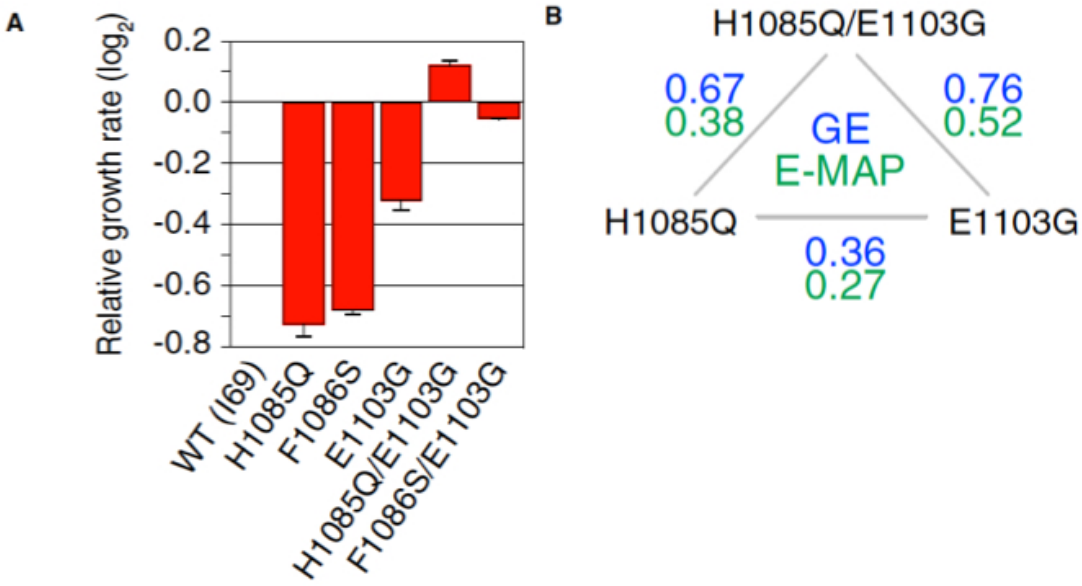


Figure 2-27. RNAPII trigger loop double and single mutant growth rates, and profile correlations between *rpb1* E1103G, *rpb1* H1085Q and *rpb1* E1103G/H1085Q. **A**. Growth rates relative to wt of *rpb1* E1103G, *rpb1* F1086S, *rpb1* H1085Q, *rpb1* E1103G/F1086S and *rpb1* E1103G/H1085Q. Means and standard deviations were derived from three replicates. **B**. pE-MAP and GE profile correlations between *rpb1* E1103G, *rpb1* H1085Q and *rpb1* E1103G/H1085Q.

Genome-Wide And Gene-Specific Effects Of Altering Polymerase Speed On In Vivo *Splicing Efficiency*

Several steps in mRNA processing are coupled to the process of transcription (Bentley, 2005). Based primarily on experiments from metazoa, a kinetic coupling model has been invoked to explain how decreasing polymerase speed affects the usage of alternative splice sites, primarily favoring the recognition of otherwise weak splice sites (de la Mata et al., 2011). We used the RNAPII mutants that differ in their *in vitro* elongation rates to determine whether this type of coupling exists in *S. cerevisiae*, where splice sites are of consensus or near-consensus sequence and the sole read-out is splicing efficiency. We used splicing-sensitive microarrays to measure the change in total, pre-mRNA, and mature mRNA for each intron-containing gene (**Figure 2-28** A; **top**) in response to mutations in the RNAPII TL. Many genes exhibit the reciprocal pre- and mature mRNA values indicative of changes in splicing (**Figure 2-28** A **side panels**) and to allow facile comparison across multiple mutants, we calculated pre mRNA/mature mRNA ratios (Clark et al., 2002) for each gene (**Figure 2-28**). Both gene-by-gene (**Figure 2-28** A) and global (**Figure 2-28** B) measures show that *rpb1* mutants characterized as fast *in vitro* (E1103G and G1097D) lead to an increase in this ratio for many genes, which indicates a defect in the splicing of those transcripts. Conversely, mutants with slow *in vitro* elongation (H1085Q and F1086S) decrease this ratio for many genes, a phenotype consistent with more efficient splicing. Thus, the global trend we observe is an anti-correlation between polymerase rate and splicing efficiency ($p<10^{-5}$ for each mutant in **Figure 2-28** A compared to WT). Using qPCR, we confirmed this

trend at several genes (**Figure 2-25** A), and found that these genes exhibit between 1% and 40% unspliced mRNA in WT, but up to 70% in G1097D (data not shown).

Given that polymerase rate and splicing efficiency are anti-correlated, it follows that the TL double mutants (*rpb1* E1103G/F1086S and E1103G/H1085Q), which transcribe at near-WT rate in vitro, should also exhibit near-WT splicing. Consistent with this prediction, these strains had very few genes with splicing defects (**Figure 2-28** B). Thus, suppressive relationships are observed in the double mutant with respect to growth, genetic and expression profiling, *in vitro* transcription rate, and also mRNA splicing. Furthermore, slowing polymerase by chemical means should phenocopy a genetically slow polymerase mutant. We therefore evaluated splicing in WT cells treated with mycophenolic acid (MPA), which is known to impede transcriptional elongation (Mason and Struhl, 2005). A 10-min MPA treatment resulted in even more genes with improved splicing when compared to a slow polymerase mutant (**Figure 2-28** B**,** $p<10^{-15}$ compared to WT), possibly because MPA elicits an acute stress. Taken together, these splicing phenotypes are consistent with a direct kinetic coupling between elongation rate and splicing *in vivo.*

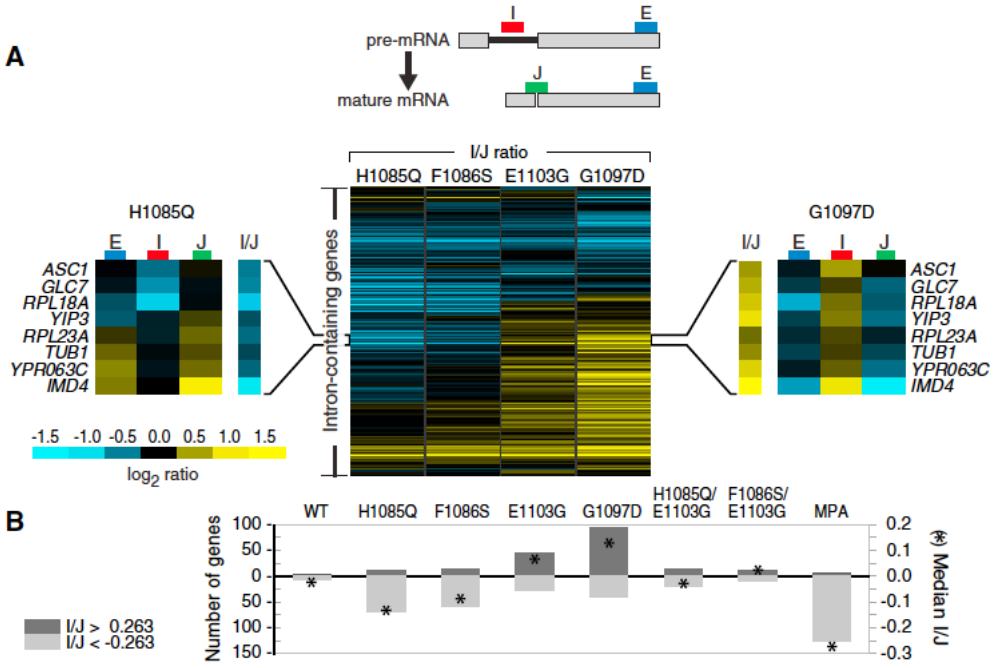


Figure 2-28. Effects of altering RNAPII transcription rate on *in vivo* splicing efficiency.

A. Top panel: microarray schematic for each intron-containing gene: probe I (intron) hybridizes to pre-mRNA, J (junction) to mature mRNA, and E (exon) to both. Center panel: heat map of I/J ratios for the slow (*rpb1* H1085Q and F1086S) and fast (*rpb1* E1103G and G1097D) mutants, corresponding to the enrichment of pre-mRNA over mature mRNA. Side panels highlight a subset of genes that behave reciprocally in fast and slow RNAPII backgrounds. **B**. Number of genes exhibiting >20% change in I/J ratio (bars, left-hand scale), and median I/J ratio (asterisks, right-hand scale) across entire array. MPA treatment was 10 minutes; WT denotes competitive hybridization between two WT cultures. See also **Figure 2-26**.

Genetic Interactions Between RNAPII Alleles And Other Mutants Reveal Relationships Between Transcription Factors And RNAPII Activity

Because of the growth suppression seen in the TL double mutants (Kaplan et al., 2012) (**Figures 2-26** C, D), we hypothesized that some of the deletion/DAmP genetic interactors we detected in our pE-MAP might directly regulate, or collaborate with, RNAPII. We reasoned that disruption of a positively acting transcription factor would result in positive genetic interactions (suppression) with fast RNAPII mutants but negative interactions (synthetic sickness) with slow mutants. Conversely, a negatively acting factor would show opposite genetic trends. To identify these factors, we sorted the deletion/DAmP mutants based on the difference in their average genetic interaction score with fast and slow RNAPII mutants (**Methods, Figure 2-29** A). We focused on genes that behaved as positively acting factors and observed that *sub1*Δ had the strongest pattern in this regard. Interestingly, previous evidence has implicated Sub1 as a positive factor in *in vitro* transcription assays (reviewed in (Conesa and Acker, 2010), as well as *in vivo* (Garcia et al., 2012). We confirmed the genetic relationships from the pE-MAP using standard growth assays, where *sub1*Δ exacerbated slow RNAPII alleles (**Figure 2-29** B) and partially suppressed fast RNAPII alleles (**Figure 2-29** C). Deletion of *SUB1* also exacerbated and suppressed the relevant RNAPII mutant phenotypes (MPA, Spt, Gal^R) (**Figure 2-30** A) (as did other mutants (**Figure 2-30** B)). Furthermore, gene expression analysis of *sub1*Δ, *rpb1* E1103G, and *sub1*Δ/E1103G showed an epistatic relationship between the E1103G mutant and *sub1*Δ (**Figure 2-29** D), consistent with

their positive interaction and suggesting that a fast RNAPII mutant can bypass the requirement for Sub1.

Our recent work has implicated changes in RNAPII activity with the alteration of start site selection *in vivo* (Kaplan et al., 2012) (**Figure 2-23**). Furthermore, Sub1 genetically interacts with TFIIB (*sua7*) (Knaus et al., 1996; Wu et al., 1999), is broadly recruited to RNAPII/III promoters *in vivo* (Rosonina et al., 2009; Tavenet et al., 2009) and was implicated as a member of the RNAPII pre-initiation complex (Sikorski et al., 2011). We therefore sought to determine whether Sub1 might modulate RNAPII start site choice. Notably, primer extension analysis at *ADH1* revealed that deleting *SUB1* led to a significant downstream shift in start site (**Figure 2-29** E**, Figure 2-30** C). Slow RNAPII TL mutant *rpb1* F1086S (**Figure 2-23** B) and *sua7* alleles (Pinto et al., 1992) (**Figure 2-25** B) also initiate downstream and are synthetic sick with *sub1*Δ. These data are consistent with the notion that Sub1 promotes transcription initiation. Double mutant analysis revealed that *sub1*Δ also exacerbated the downstream start site shift of the slow RNAPII TL allele, *rpb1* F1086S, and slightly suppressed the *rpb1* E1103G allele (**Figure 2-29** E). Because *sub1*Δ has also been linked to another RNA processing step, namely 3′ end processing, we examined the effect of *sub1*Δ on splicing and observed a statistically significant increase in splicing efficiency ($p<10^{-8}$), again phenocopying the slow RNAPII mutants (**Figure 2-29** F).

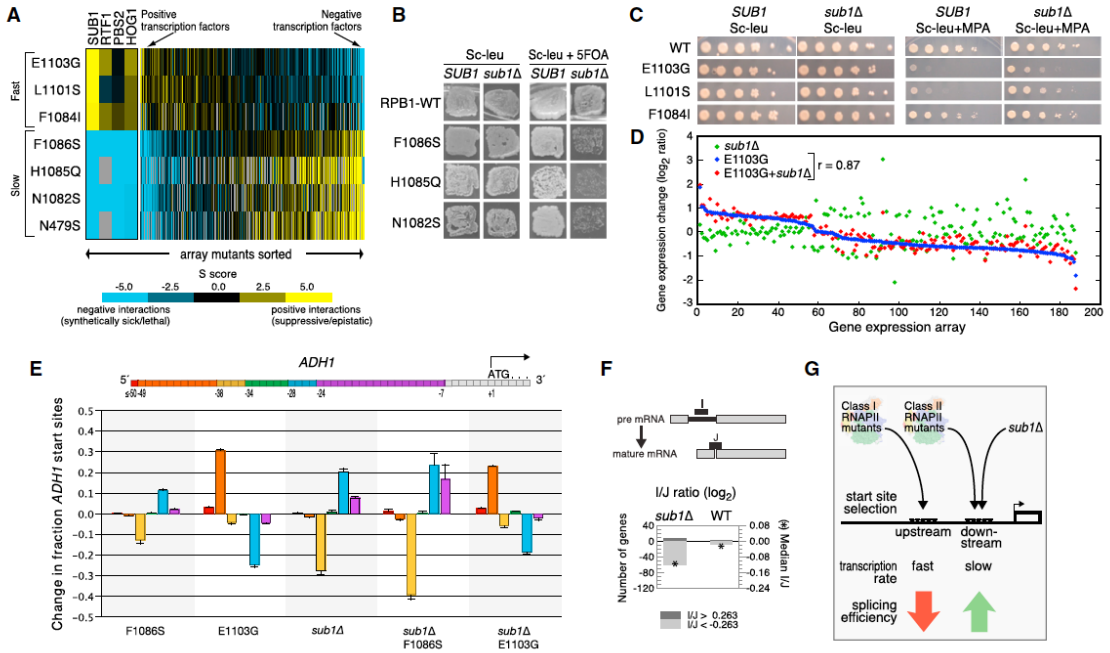
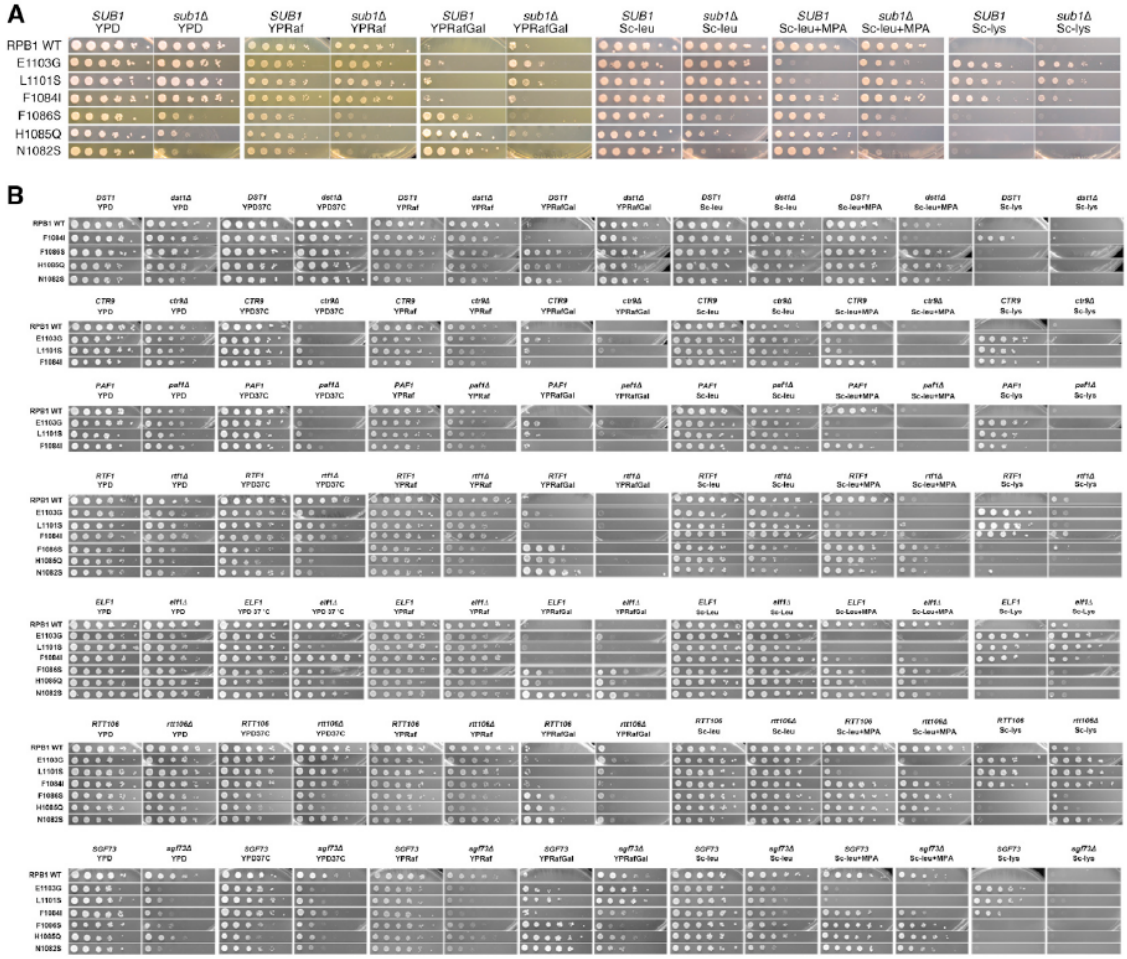


Figure 2-29. Genetic interaction patterns with fast and slow RNAPII mutants reveal Sub1 as a transcription factor that regulates start site selection and influences mRNA splicing.

A. Genetic profiles of library mutants, sorted on the difference between their average interaction with fast and slow RNAPII mutants. **B**. Patch tests examining the sensitivity of slow TL mutants to *sub1*Δ. WT *RPB1* plasmid covering *rpb1* mutants in left panel is lost in right panel. **C**. Spot tests examining the effect of *sub1*Δ on fast mutants in absence (left) and presence (right) of MPA. **D**. Comparison of *sub1*Δ effect on gene expression in E1103G (difference between red and blue) and WT (difference between green and y=0). Included are all array transcripts exhibiting a >1.5-fold expression change in at least one of the three mutants. Transcripts are sorted by expression change in E1103G. **E**. Primer extension at *ADH1* to map transcription start sites for F1086S, E1103G and s*ub1*Δ (**Figure 2-30** C). Bar colors correspond to sequence windows in the *ADH1* schematic (top) and heights specify the fraction change of transcription start in mutant compared to WT. Error bars represent standard deviations. **F**. Splicing microarray analysis of *sub1*Δ, as in **Figure 2-28**. Number of genes exhibiting a >20% change in I/J ratio (bars, left-hand scale), and median I/J ratio (asterisks, right-hand scale). **G**. Model for the effect of Sub1 and RNAPII activity on start site selection and splicing. Fast RNAPII mutations (Class I) result in upstream transcription start and diminished splicing efficiency, whereas *sub1*Δ or slow RNAPII mutations (Class II) shift transcription start downstream and enhance splicing.

Figure 2-30. Spot tests to determine Gal^R, MPA sensitivity, and Spt$^-$ phenotypes of *sub1*Δ or other deletion mutants with fast and slow RNAPII alleles and primer extension of *sub1*Δ mutants.

A. Spot tests on relevant media were carried out to determine the effects of s*ub1*Δ on the transcription-related phenotypes of fast and slow RNAPII alleles, as described in **Figure 2-17** and **Methods**. **B**. Spot tests on relevant media were carried out to determine the effects of deleting known or predicted transcription factors on the transcription-related phenotypes of fast and slow RNAPII alleles, as described in **Figure 2-17** and **Methods**. **C.** Sequencing gel for primer extension analysis at *ADH1* to identify the effects of s*ub1*Δ on start-site selection. Lanes 1-4 correspond to Sanger dideoxy sequencing reactions as reference ladders and the following lanes carry reverse transcription products from RNAPII mutants in *SUB1* and *sub1Δ* background as indicated.


A
SUB1 YPD
sub1Δ YPD
SUB1 YPRaf
sub1Δ YPRaf
SUB1 YPRafGal
sub1Δ YPRafGal
SUB1 Sc-leu
sub1Δ Sc-leu
SUB1 Sc-leu+MPA
sub1Δ Sc-leu+MPA
SUB1 Sc-lys
sub1Δ Sc-lys
RPB1 WT
E1103G
L1101S
F1084I
F1086S
H1085Q
N1082S
B
C
G A T C
RPB1 WT
F1086S
E1103G
+ Δ
SUB1
-78
-62
-50
-37
-27
-24
-20
-16
-7

Based on these data, we propose a model where transcription start and splicing are intimately coupled with RNAPII elongation: fast RNAPII mutations result in upstream transcription start and diminished splicing, whereas slow mutations or *sub1*Δ give rise to downstream transcription start and enhanced splicing (**Figure 2-29** G). Given the possibility of direct coupling between start site selection and downstream mRNA processing, we globally measured splicing defects in *sua7*-3 (TFIIB) (Pinto et al., 1994; Wu et al., 1999) and *tgf2*Δ261-273 (TFIIF) (Eichner et al., 2010), mutants that exhibit downstream and upstream start site selection, respectively (**Figure 2-25** B). If splicing were strictly coupled to start site choice, one would expect these mutants to have similar splicing defects to the slow and fast RNAPII mutants. However, such a correlation was not observed (**Figure 2-25** C), suggesting these processes are, in fact, genetically separable. Taken together, these data support a model in which the catalytic rate of RNAPII has multiple, separable effects on start site selection and mRNA processing, and highlight the importance of WT elongation rate for multiple steps in gene expression.

Discussion

In this study, we have described an important extension of our E-MAP genetic interaction mapping strategy, the functional interrogation of individual protein residues. We have first used this approach to genetically dissect RNAPII and demonstrate how the pE-MAP successfully characterized functionally distinct, individual amino acids within this multi-functional complex. This analysis not only provided insight into global

structure-function relationships within RNAPII but also specific details about how RNAPII regulates, and is regulated by, different factors and processes.

Insights Into RNAPII Function Derived From The pE-Map

Multiple aspects of the genetic interaction map provided insights into transcriptional regulation. First, by comparing the RNAPII genetic interaction profiles with those from previous deletion/knockdown studies, we could assign function to individual residues, which allowed us to generate a point mutant-protein complex connectivity map (**Figure 2-21**, **Figure 2-22**). Furthermore, analyzing double mutants, both within RNAPII itself (**Figure 2-26** C, D) and between RNAPII and other genes (**Figure 2-29**), allowed for a better understanding of TL function, as well as the identification of other factors that directly or indirectly impinge on RNAPII activity. Indeed, the pE-MAP allowed us to identify putative transcription factors (negative and positive), such as Sub1. While Sub1 was previously implicated at the promoter (Sikorski et al., 2011), we here demonstrated that it positively regulates RNAPII by influencing start site usage (**Figure 2-29**). The epistatic relationships between fast RNAPII alleles and *sub1*Δ suggest that Sub1 activity may be bypassed when RNAPII catalytic activity is increased. Collectively, our data indicate that Sub1 plays a direct role in transcriptional initiation and influences mRNA splicing, possibly via its effect on elongation (Garcia et al., 2012). Overall, the genetic interactions with TL mutants should provide a rich dataset of potential positive and negative transcription factors for further study.

Finally, structural analysis revealed that mutations close in 3-D space have very similar genetic profiles, including those in different subunits (**Figure 2-20** B), suggesting

that structural information is ultimately contained within the p-EMAP and can be used to identify specific protein-protein interaction interfaces. For example, the Rpb2 mutation E437G/F442S, which shifts start site selection upstream, is in a domain that contacts TFIIF at a region in which mutations also result in upstream start site shifts (Chen et al., 2007; Eichner et al., 2010). Furthermore, the identification of an Rpb7 mutant (D166G) that alters start site selection is intriguing, as this region interacts with TFIIH in a cryo-EM structure of the RNAPII pre-initiation complex (He et al., 2013) and TFIIH also alters start site selection (Goel et al., 2012). Therefore, when this analysis is extended to other molecular machines, the resolving power of the genetic profiles can be harnessed, with other biophysical techniques such as cross-linking or electron microscopy, to help identify the specific regions of proteins that physically interact.

Coordination Of Transcriptional Rate With Start Site Selection And mRNA Splicing

The importance of maintaining WT rates of transcription is evidenced by the phenotypic defects observed in the fast and slow mutants (Kaplan et al., 2012) (**Figure 2-26**, **Figure 2-28**), as well as the striking mutual suppression seen when combining two mutations that individually make RNAPII slow or fast (**Figure 2-26**, **Figure 2-28**). Interestingly, the pE-MAP identified two groups of RNAPII mutants: one that preferentially initiates upstream and exhibits an increased rate of transcription, and one that initiates downstream and transcribes slowly (**Figure 2-29**) (Kaplan et al., 2012). We propose that both phenotypes are direct consequences of the efficiency of nucleotide selection and incorporation, since the addition of the first nucleotides at initiation is biochemically similar to adding nucleotides during elongation. These data support the

model that RNAPII engages in "scanning" during initiation in *S. cerevisiae* (Giardina and Lis, 1993; Kaplan, 2013; Kaplan et al., 2012; Kuehner and Brow, 2006). Whether RNAPII catalysis drives this scanning, or whether scanning occurs in the absence of nucleotide incorporation, perhaps driven by TFIIH, is unknown. We also note that because changes in start site selection alter the 5′UTR length and composition, initial transcription decisions may have downstream effects on gene expression such as changes in RNA stability or translational efficiency of the mRNA (Arribere and Gilbert, 2013; Rojas-Duran and Gilbert, 2012).

The pE-MAP has allowed important insights into the co-transcriptional process of mRNA splicing. It is now clear that most introns are removed while RNA polymerase is still associated with the DNA template. In metazoans, alternative splicing decisions can be influenced by factors impinging on transcription, including promoter identity and polymerase speed (reviewed in (Perales and Bentley, 2009)). Slowing the rate of elongation by mutation of RNAPII or chemical means can improve the recognition of splice sites that deviate from consensus signals (de la Mata et al., 2003; Howe et al., 2003; Ip et al., 2011). Because the spliceosome undergoes stepwise assembly on each intron, slowing transcription can afford more time for formation of the catalytically active machine before transcription of a downstream, stronger site. While budding yeast lack alternative splicing, it nonetheless follows that the efficiency of co-transcriptional splicing would be favored by allowing sufficient time for spliceosome assembly. Indeed, recent work suggests that RNA polymerase may slow down to favor co- vs. post-transcriptional splicing (Aitken et al., 2011; Alexander et al., 2010; Carrillo Oesterreich

et al., 2010). Our microarray analyses show a clear trend in which splicing efficiency is generally anti-correlated with transcription rate (**Figure 2-28**). This is the first genome-wide analysis to interrogate kinetic coupling in *S. cerevisiae*. Moreover, this is the first instance in which faster and slower polymerases have been directly compared; thus, this work satisfies the previously untested prediction that rate and splicing efficiency would be generally anti-correlated.

Interestingly, we observed correlation between start site selection, elongation rate and splicing efficiency. In fact, promoter-proximal events are known to be able to influence downstream RNA transactions: promoter identity can influence alternative splicing or mRNA stability in other systems (Cramer et al., 1997; Harel-Sharvit et al., 2010; Trcek et al., 2011), and 5′UTR length, determined by start site selection, can strongly alter translation efficiency in budding yeast (Rojas-Duran and Gilbert, 2012). However, when we measured splicing efficiency using mutants in the general transcription factors TFIIF and TFIIB that alter start site selection, we found that not all initiation phenotypes are predictive of splicing efficiency (**Figure 2-25** C). This suggests that RNAPII catalytic rate has several, separable effects on gene expression, a claim supported by recent evidence showing kinetic coupling between RNAPII transcription and Sen1-dependent termination (Hazelbaker et al., 2013).

Taken together, our data highlight the important impact of transcription speed determined by the genetic status of RNAPII and trans-acting factors (e.g. Sub1) on start site selection and mRNA splicing. We propose that RNA polymerase may have been evolutionarily tuned to coordinate between multiple steps in gene expression, and we

predict that polymerase rate may influence multiple additional co-transcriptional steps in gene expression, including mRNP assembly, 3′ end processing, and export.

Future Studies Using The pE-Map Approach

S. cerevisiae RNAPII provided the groundwork for validating the pE-MAP approach. Since we learned so much from an already well-studied machine, it suggests that this technology should be applied to other molecular machines, including those involved in translation (e.g. ribosome), protein quality control (e.g. proteasome, HSP70, HSP90), chromatin (histones) and DNA replication (DNA polymerases). This analysis can also be carried out in other organisms that are genetically tractable and where high-throughput genetic interaction mapping methodologies have been created, including *S. pombe* (Roguev et al., 2008; Roguev et al., 2007; Ryan et al., 2012) and *E. coli* (Butland et al., 2008; Typas et al., 2008). Furthermore, the platform could also be used to gain structural insight into proteins and complexes that have not yet been structurally characterized. Finally, as genetic interaction mapping strategies become more prevalent in mammalian cells (Bassik et al., 2013; Laufer et al., 2013; Lin et al., 2012; Roguev et al., 2013) and with the development of genome editing using various double-strand break technologies (Gaj et al., 2013), similar work characterizing the function of individual amino acids will have great impact on understanding how point mutations in specific genes result in different disease states.

Methods

E-MAP-compatible *MAT*a RNAPII mutant strains carrying a marked *rpb* deletion and mutant *rpb* on a *CEN* plasmid were mated with 1200 *MAT***a** DAmP/deletion strains by pinning on solid media. Sporulation was induced and double mutant *MAT***a** spores were isolated on selective media. Genetic interactions were scored based on double mutant colony sizes, which were extracted using automated imaging software.

For gene expression and splicing arrays, total RNA was extracted from mutant and WT log-phase cells. Competitive hybridizations were performed between mutant and WT cDNA (splicing) or cRNA (gene expression).

CTF assays were carried out by plating strains carrying *ade2-101* and a chromosome VII fragment containing *SUP11* on SC medium with 20% adenine and measuring the fraction of red colonies. The red color caused by *ade2-101* is counteracted by *SUP11*; chromosome fragment loss results in red colonies.

For transcription start site analysis, we performed primer extension from a ^{32}P end-labeled oligo annealed to total RNA. cDNAs were separated by PAGE and bands quantified.

Plate Phenotyping

Plate phenotyping of RNAPII mutants was done as described (Kaplan et al., 2012). Briefly, *CEN LEU2* plasmids containing a mutant *rpb* RNAPII subunit gene were transformed into an appropriate Leu$^-$ strain with corresponding endogenous RNAPII subunit gene knocked out but complemented with a *CEN URA3* WT *RPB* subunit gene. Leu$^+$ transformants were patched on solid medium lacking leucine and replica-plated to

medium lacking leucine but containing 5-fluoroorotic acid (5-FOA, Gold Biotechnology) to select against cells maintaining *RPB* WT *URA3* plasmids. *sub1Δ* + RNAPII mutant strains for direct testing of double mutant phenotypes were constructed based on CKY283 and analyzed in the same fashion as RNAPII single mutants. Yeast media used in phenotyping were prepared as previously described (Amberg et al., 2005; Kaplan et al., 2012). YPD medium contained yeast extract (1% w/v final, BD), peptone (2% w/v final, BD), dextrose (2% final w/v), bacto agar (2%, BD) and supplemental adenine and tryptophan. Alternate carbon source YP media were YP raffinose (2% final w/v, USB) and YP raffinose (2% final w/v, USB) plus galactose (1% final w/v), with plates also containing antimycin A (1µg/ml, Sigma-Aldrich). Synthetic complete medium lacking leucine (SC-Leu) was as described (Kaplan et al., 2012). Mycophenolic acid (MPA, Sigma-Aldrich) was added to SC-Leu at 20µg/ml final concentration from a 10mg/ml stock in ethanol.

Galactose sensitivity or resistance (Gal^r or Gal^s, determined based on strain grown on YP raffinose plus galactose) phenotypes are based on an allele of *GAL10*, *gal10Δ56*, that has a deletion of the major *GAL10* polyadenylation signal (Greger et al., 2000; Greger and Proudfoot, 1998; Kaplan et al., 2005). A consequence of this deletion is that *GAL10* transcription reads through into the downstream *GAL7* promoter, interfering with *GAL7* expression. Under conditions where other Gal genes are expressed, specifically Gal1p, Gal7 protein is required to prevent galactose toxicity caused by Galactose catabolism intermediate, thus defects in *GAL7* due to transcriptional interference from

gal10Δ56 cause the Gal^s phenotype, and modulation of this transcription defect can lead to Gal^r or galactose super-sensitive, Gal^{ss} phenotypes.

The Spt phenotype studied here relates to the *LYS2* allele, *lys2-128∂* (Simchen et al., 1984), and suppression of the lysine auxotrophy (Lys$^-$ phenotype) conferred by a Ty *∂* element insertion into the 5′ end of the *LYS2* gene present in this allele. The suppression of Lys$^-$ phenotype (growth on medium lacking lysine) is referred to as the Spt$^-$.

RNAPII Subunit Mutagenesis And Site-Directed Mutagenesis

PCR-based random mutagenesis was performed to introduce mutations into RNAPII subunits. For each individual RNAPII subunit, one or more screens were performed, mutagenizing different parts of longer subunit genes in different screens (*RPO21/RPB1* or *RPB2*), or the entire open reading frame (*RPB3*, *RPB7*, *RPB11*) for smaller subunit genes. Plasmids containing individual subunits and flanking DNA allowing for expression and termination from native elements were constructed in pRS-based *CEN*, low copy vectors (Sikorski and Hieter, 1989) for complementation of appropriate RNAPII subunit gene deletions. Mutagenesis conditions were under standard PCR conditions (1.5 mM $MgCl_2$, 200 μM dNTPs) where 6 individual reactions for each mutagenized region were performed for 35 cycles of amplification followed by seeding of 6 additional reactions for 35 more cycles, followed by pooling of reactions for each mutagenized region. Mutagenized PCR products were transformed into yeast along with appropriate gapped plasmid vector lacking most of the region to be mutagenized. Selection of vector marker (*LEU2*) after transformation allowed plasmids putatively

created through gap repair recombination between gapped-vector and mutagenized PCR products to be obtained. These transformant populations represented putative mutant libraries that were screened for mutant phenotypes upon standard plasmid shuffling (Boeke et al., 1987) to remove the wild type *RPB* plasmid, allowing the transformed plasmid to be the only copy of the particular *RPB* subunit in the colony. Following replica-plating to 5FOA, plates were additionally replicated to a number of other media for phenotyping. Candidate mutants were streaked and rescreened for phenotypes followed by either dominance/recessive/plasmid linkage tests or direct plasmid isolation and retransformation to assay plasmid linkage/reproducibility of observed phenotypes. Mutant plasmids were isolated from yeast by standard procedures and portions of insert that had been subjected to PCR in initial mutagenesis were sequenced to identify alterations.

For *RPB3,* two independent replicate screens were performed, *Hin*dIII sites present in flanking genomic DNA were used to remove wild type sequence and gap-repair mutagenize the entirety of the *RPB3* ORF in an individual screen. For *RPB7*, vector *Xho*I and *Sac*I sites were used to remove the entire *RPB7* insert to allow mutagenesis of the entire *RPB7* ORF. For *RPB11*, a vector *Xho*I site and a *Hin*dIII site present in the 3′ end of *RPB11* were used to remove wild type *RPB11* and mutagenize the majority if not all of the *RPB11* ORF. For *RPB2*, three individual screens were performed: the first utilized *Nde*I and *Hin*dIII sites to remove ~700 basepairs of RPB2 ORF 5′ sequence for mutagenesis, the second utilized *Hin*dIII and *Msc*I to remove a 1.6 kb internal ORF fragment for mutagenesis, and the third used an internal *Msc*I and vector *Sac*I site to

remove a 1.6 kb *RPB2* ORF 3′ sequence for mutagenesis. For *RPO21/RPB1*, three screens were performed. For the first, engineered silent *Bam*HI and *Pst*I restriction sites were utilized for removal of an internal *RPO21/RPB1* ORF fragment for mutagenesis. For the second, a 3′ *RPO21/RPB1* ORF fragment up to the CTD was removed with digestion by *Xba*I and *Bsi*WI for mutagenesis. For the third, the entire *RPO21/RPB1* ORF 3′ region was removed using an *Xba*I-*Sna*BI for mutagenesis. In general, 3000-5000 colonies were screened per independent screen and mutagenesis levels were approximated by number of 5FOA-sensitive colonies (representing lethal mutations and were between 2-7% in different screens).

For creation of *URA3*-marked *CEN* plasmids for pE-MAP screens, RNAPII alleles were moved from *LEU2*-marked *CEN* versions into pRS316 or pRS416 by standard cloning procedures. *URA3*-marked RNAPII mutant plasmids were then transformed into pE-MAP *rpb*Δ strains and screened for loss of wild type *RPB LEU2* plasmids.

E-MAP Analysis

E-MAP amenable strains for RNAPII mutants were created by transformation of MATa spore selectable marker (*mfa1p::HIS3*) into relevant MATa *rpb*Δ (complemented with WT *RPB CEN LEU2* plasmid) strains via recombination into the *CAN1* locus, generating *mfa1p::HIS3::can1*Δ. Mating type was switched to *MAT*a following standard procedures and mutant *rpb* RNAPII subunit genes were introduced on *CEN URA3* plasmids, followed by screening for loss of WT *RPB CEN LEU2* plasmid. E-MAP experiments and scoring of genetic interactions were then carried out as previously described (Collins et al., 2007b; Collins et al., 2006; Schuldiner et al., 2006).

ROC Curves

All library deletion mutants that exist in both the RNAPII pE-MAP and the previously published deletion E-MAP (Collins et al., 2007b) were extracted, resulting in 404 mutants used for this analysis. First, pairs of proteins encoded by these genes were defined as physically interacting if listed with PE scores >2 and non-interacting if not (Collins et al., 2007a). Next, Pearson correlation coefficients were computed for all pairwise combinations of the 404 mutants, based on their pE-MAP profiles. An ROC curve was then generated to determine the power of these correlations to predict physical interactions between encoded proteins. Prior to generating the reference ROC curve, deletion E-MAP query strains with more missing data than the sparsest RNAPII mutant were removed, as were those that also exist in the library set. From this filtered E-MAP, 53 rows were randomly selected 1000 times, and an ROC curve was generated for each run. The median AROC and corresponding curve are reported in **Figure 2-20**.

Genome-Wide Gene Expression

Strains were streaked from -80°C stocks onto plates and grown for 3 days. Liquid cultures were inoculated with independent colonies and grown overnight in Synthetic Complete (SC) medium: 2g/l Drop out mix Complete and 6.71g/l Yeast Nitrogen Base without AA, Carbohydrate & w/AS (YNB) from US Biological (Swampscott, USA) with 2% D-glucose. Overnight cultures were diluted to OD_{600}=0.15 in 1.5ml fresh medium and grown at 30°C in a 24 well plate in a Tecan Infinite F200 under continuous shaking. Growth curves were made for the mutant cultures (two cultures from two isolates) as well as for two wt inoculates, grown in parallel. Mutant and wt cells were

harvested by centrifugation (6100 rpm, 3 min) at mid-log phase at OD_{600}=0.6, and pellets were immediately frozen in liquid nitrogen after removal of supernatant. Up to eleven mutant strains could be grown on a single day. wt cultures were grown parallel to the deletion mutants to assess day-to-day variance. RNA isolation and purification was performed (Lenstra et al., 2011). In short, total RNA was prepared by phenol extraction and cleaned up using a customized Sciclone ALH 3000 Workstation. For each sample, external control poly-A+ RNAs were added in equimolar amounts to the total RNA to enable monitoring of global changes in mutants (van de Peppel et al., 2003). For each micro-array analysis, 1.25μg labeled sample cRNA and 1.25μg reference cRNA was hybridized per slide. Each hybridization performed within this project was subjected to a number of quality controls. Some of these are based on the data from one single hybridization, while others are based on comparing data from one single hybridization against the wt grown in parallel (Lenstra et al., 2011). Two channel microarrays were used. RNA isolated from a large amount of wt yeast from a single culture was used as a common reference. This common reference was used in one of the channels for each hybridization and used in the statistical analysis to obtain an average expression-profile for each mutant relative to the wt. Two independent cultures were hybridized on two separate microarrays. Each gene is represented twice on the microarray, resulting in four measurements per mutant.

Correlations

All correlations are Pearson correlation coefficients, unless otherwise noted. Genetic interaction correlations are based on the complete genetic interaction profiles, and gene expression correlations on the genome-wide expression profiles.

Identification Of Functional Links Between RNAPII Mutants And Protein Complexes

Published genetic interaction data (Costanzo et al., 2010) were scaled to E-MAP format using a non-linear scaling method (Ryan et al., 2012), and processed to remove duplicate entries of genes represented by more than one allele (deletion, temperature sensitive or hypomorphic). If the deletion was present it was kept and all other alleles removed, and if no deletion existed the allele to keep was selected randomly. Pearson correlation coefficients were computed for each RNAPII mutant against all genes in this dataset, based on the genetic interaction profiles with the intersecting library mutants. The genes were then grouped by membership in physical complexes of their encoded proteins. Complex definitions were collected from (Benschop et al., 2010), and modified by first removing all binary complexes, and then excluding any gene assigned to more than one complex. To identify allele-complex pairs which were significantly correlated, we used a one-sided Mann-Whitney U test to compare the correlations between each RNAPII mutant and the members of each complex to (i) the correlations between the same mutant and all genes not in that complex, and to (ii) the correlations between the same complex and all other mutants. The highest p-value of the comparison to (i) or (ii) was recorded. False discovery rates (FDR) were computed using the method of Benjamini and Hochberg (Benjamini and Hochberg, 1995).

Chromosome Transmission Fidelity (CTF) Assay

RNA polymerase mutants containing a chromosome VII fragment and *ade2-101* were generated. Red colony color is caused by accumulation of pigment due to a block in adenine production caused by the *ade2-101* (ochre) mutation. The block is relieved in the presence of the *SUP11* gene on the telocentric arm of the chromosome fragment, encoding an ochre-suppressing $tRNA^{Tyr}$. Cells that lose the chromosome VII fragment develop red color so that colonies exhibiting unstable inheritance of the chromosome fragment develop red sectors. Mutant strains were plated at 100-400 cells a plate on synthetic complete medium with only 20% adenine and left to grow into colonies for 7 days at 30°C. Replicates were counted to obtain the percentage of sectored colonies over total colonies for each mutant strain (Spencer et al., 1990; Yuen et al., 2007).

Primer Extension

Primer extension assays were performed essentially as previously described (Ranish and Hahn, 1991). Briefly, 30µg total yeast RNA purified as described (Schmitt et al., 1990) was annealed with ^{32}P end-labeled oligo priming downstream of start sites in a 15µl reaction volume instead of 10µl, followed by reverse transcription reaction supplemented with RNase Inbibitor (Fermentas) by M-MLV Reverse Transcriptase (Fermentas), in a total volume of 45µl instead of 30µl. Products were separated on 8% polyacrylamide gels (19:1 acrylamide:bisacrylamide) containing 1XTBE and 7M urea. Reported values are means derived from at least three independent experiments.

Growth Rate Assay

Strains were cultured overnight in YPD, then resuspended to $OD_{600} = 0.3$ in the morning, and grown at 30°C until reaching $OD_{600} = 0.7 – 1.0$. These cultures were diluted in YPD to $OD_{600} = 0.05$, 100μl/well, in a Corning Costar 3631 clear bottom 96-well plate. Each culture was distributed into 3 wells as technical replicates. Plates were sealed with Corning 3930 clear polysterene lids, and edges taped to avoid evaporation. Growth curves were measured at 30°C in a Tecan Infinite 200 PRO microplate reader, controlled by Tecan Magellan software. OD_{600} measurements were performed every 15 minutes (top reading, 25 reads per well), with orbital shaking during incubation. Curves were fitted to the measured time series with a model-free spline method from the grofit package (Kahm et al., 2010), in the statistical computing environment R. Exponential growth rates were determined by grofit as the maximum slopes of the spline fits. For each set of technical replicates, the $\log_2$ ratio of mutant to wt growth rate was computed. Reported averages and standard deviations were calculated over the $\log_2$ ratios in the three replicate sets.

Splicing Microarray Assay

Cultures were grown in rich medium according to standard techniques (Guthrie and Fink, 2002). Saturated cultures were diluted to $OD_{600} = 0.1$ in the morning and allowed to grow at 30°C until reaching mid-log phase (OD_{600}=0.5-0.7). Mutant strains and an isogenic wild-type were collected by centrifugation and snap frozen in liquid nitrogen. Where indicated, a wild-type log-phase culture was treated with mycophenolic acid (30μg/ml) for 10 minutes, harvested using vacuum filtration and snap frozen in

liquid nitrogen. Total cellular RNA was isolated using a hot acid phenol extraction generally as outlined in (Schmitt et al., 1990), but with some specific modifications (Bergkessel et al., 2011). cDNA from each strain was synthesized, and labeled with Cy3 or Cy5 according to the low-throughput sample preparation method (Pleiss et al., 2007). The optimized oligos (Pleiss et al., 2007) were robotically arrayed onto poly-L-lysine coated glass slides (slides from ThermoScientific C40-5257-M20) and slides were processed according to (DeRisi et al., 1997; Pleiss et al., 2007). Microarrays were scanned using Axon Instruments GenePix 4000B at 635nm and 532nm wavelengths and image analysis was done using Axon Instruments GenePix Pro version 5.1. Spots were manually removed from analysis if they contained uncharacteristically high background or obvious defects; the ratio of the median intensity for 535nm and 625nm was calculated for each remaining spot. Each biological replicate contains 6 technical replicates for each feature as well as dye-flipped replicates, which were combined and normalized (Pleiss et al., 2007). For all mutants, data from at least two biological replicates were used. To emulate a classic splicing measure that compares the levels of pre-mRNA and mature mRNA (Pikielny and Rosbash, 1985), we computed I/J ratios as $\log_2(I/J) = \log_2(Intron_{mutant}/Intron_{WT})-\log_2(Junction_{mutant}/Junction_{WT})$ for each gene. The reported p-values were calculated via one-sided Wilcoxon signed rank tests, comparing each I/J distribution to that resulting from direct hybridization of cDNA from two separately grown wild-type cultures.

qPCR Assay For Splicing Efficiency

Splicing efficiency was measured by qPCR of cDNA as in (Pleiss et al., 2007) with modifications as described below. Ten µg of RNA were treated with 4U RQ1 DNaseI (Promega) in 16µL according to manufacturer instruction (20 min 37˚C). 4µL Stop Solution was added, and DNase was inactivated by incubation at 65˚C for 10 min. Samples were primed with random 9-mers in 40µL (50mM Tris-HCl pH8.3, 75mM KCl, 3mM $MgCl_2$, 10mM DTT, 400ng dN9; 5 min 65˚C; 5 min on ice) and divided into equal 40µL reverse transcription reactions with M-MLV reverse transcriptase (+RT or -RT in 50mM Tris-HCl pH8.3, 75mM KCl, 3mM $MgCl_2$, 10mM DTT, 250mM each dNTP; 2 hours 42˚C). qPCRs were performed in 25µL consisting of 1X Standard *Taq* Buffer and 1.25U *Taq* (New England Biolabs), 200mM each dNTP, 0.0013% SYBR Green I (Sigma), 600nM each primer, DMSO as needed, and 25x – 2500x dilution of cDNA (-RT samples diluted equivalently to match +RT). Amplification reactions were performed on a C1000 ThermoCycler (BioRad) (94˚C 3 min and 35 cycles of 94˚C 15 sec, 50˚C 15 sec, 72˚C 15 sec, plate read). Each qPCR run was finished with a melt curve to confirm homogeneity of the amplified product. Starting quantity was calculated from a genomic DNA standard curve for each primer set; standard curve reactions not falling in the linear range were removed manually. Two technical replicates were performed for 2-8 biological replicates (F1086S n=8, H1085Q n=5, sub1Δ n=2, MPA n=3, E1103G n=6, G1097D n=5). -RT control samples yielded negligible amplification (data not shown). Each gene was measured using primer sets specific for pre-mRNA and total mRNA. To generate the graph in **Figure 2-25**, the pre-mRNA/total mRNA ratio for each mutant was

normalized to a within-experiment wild-type before taking the mean of the biological replicates.

Sorting Array Mutants On Their Interactions With Fast And Slow Mutants

Mean genetic interaction scores with the fast or slow RNAPII groups were calculated for all array mutants with at least two scores in each group. The 404 mutants with different signs of mean (fast) and mean(slow) were then sorted by the difference of the means.

CHAPTER III

RELATIONSHIPS OF RNA POLYMERASE II GENETIC INTERACTORS TO TRANSCRIPTION START SITE USAGE DEFECTS AND GROWTH IN *SACCHAROMYCES CEREVISIAE*[*]

Disclaimer

Chapter III is a reprint of a publication for which I am first author. In order to determine the relationship between Pol II mutants' start site defects and general transcription factor (GTF) functions, I combined Pol II start site mutants with GTF start site mutants. These experiments were to determine if GTF start site mutant enhancement or suppression of Pol II mutant *start site* defects correlated with enhancement or suppression of Pol II mutant *growth* defects. The summary section of this chapter is the abstract of the publication, and the rest is as published. All of the experimental work in this chapter was done by myself.

[*]Reprinted with permission from "Relationships of RNA polymerase II genetic interactors to transcription start site usage defects and growth in *Saccharomyces cerevisiae*" by Jin and Kaplan G3 January 1, 2015 vol. 5 no. 121-33 Copyright © 2015 Jin and Kaplan

Summary

Transcription initiation by RNA Polymerase II (Pol II) is an essential step in gene expression and regulation in all organisms. Initiation requires a great number of factors, and defects in this process can be apparent in the form of altered transcription start site (TSS) selection in Saccharomyces cerevisiae (Baker's yeast). It has been shown previously that TSS selection in S. cerevisiae is altered in Pol II catalytic mutants defective in a conserved active site feature known as the trigger loop. Pol II trigger loop mutants show growth phenotypes in vivo that correlate with biochemical defects in vitro and exhibit wide-ranging genetic interactions. We assessed how Pol II mutant growth phenotypes and TSS selection in vivo are modified by Pol II genetic interactors in order to estimate the relationship between altered TSS selection in vivo and organismal fitness of Pol II mutants. We examined whether the magnitude of TSS selection defects could be correlated with Pol II mutant-transcription factor double mutant phenotypes. We observed broad genetic interactions among Pol II trigger loop mutants and General Transcription Factor (GTF) alleles, with reduced activity Pol II mutants especially sensitive to defects in TFIIB. However, Pol II mutant growth defects could be uncoupled from TSS selection defects in some Pol II allele-GTF allele double mutants, while a number of other Pol II genetic interactors did not influence ADH1 start site selection alone or in combination with Pol II mutants. Initiation defects are likely only partially responsible for Pol II allele growth phenotypes, with some Pol II genetic interactors able to exacerbate Pol II mutant growth defects while leaving initiation at a model TSS selection promoter unaffected.

Introduction

Pol II is essential for expression of all protein-coding genes, and determining how the combined defects of Pol II activity mutants in all steps of the Pol II cycle (*i.e.* initiation, elongation, termination, cotranscriptional events) lead to growth defects is a difficult task. Initiation, the first step in transcription, is highly conserved with regulation requiring a great number of factors (Cramer et al., 2008; Hahn, 2004). Classical biochemical experiments using model promoters have shown that Pol II requires general transcription factor (GTF) TFIID, TFIIB, TFIIF, TFIIE, TFIIH for promoter recognition, formation of the preinitiation complex (PIC), promoter melting, and transcription start site (TSS) selection. The integration of these and other factors determines the efficiency of any particular promoter and the sequences that will be used to initiate transcription. Much remains to be understood about the functions of GTFs and how they integrate with Pol II activity during initiation. The effects of Pol II activity and roles of GTFs in initiation are especially visible in the process of TSS selection in *S. cerevisiae*, where most promoters use multiple start sites, and this usage is sensitive to a number of factors. Therefore, examination of mutant effects on TSS selection allows a window on the initiation process *in vivo*.

The most well-known core promoter element, the TATA box (consensus TATAWAWR motif in *S. cerevisiae*), is highly conserved throughout evolution but is only found at a subset of promoters. TSS selection at TATA element-dependent promoters in *S. cerevisiae* involves recognition of transcription start site positioned 40−120 nucleotides (nt) downstream from the TATA box and the use of multiple start sites

at most promoters, whether classified as TATA-containing (contains consensus TATA box) or not (Basehoar et al., 2004; Corden, 2008; Dvir, 2002; Li et al., 1994)(Jin and Kaplan, unpublished observations). Such extensive downstream positioning of TSS in yeast is distinct from other eukaryotes for TATA-dependent promoters, where starts are more tightly focused ~30 nt downstream of the beginning of the TATA box. Despite this difference in TSS distance to promoter element, in *S. cerevisiae* promoter melting appears to start 20-30 nt downstream of the TATA box, and thus is similar to higher eukaryotes even though start sites can be over 100nt further downstream in *S. cerevisiae* (Giardina and Lis, 1993). These results suggested that *S. cerevisiae* Pol II scans for favorable TSS subsequent to promoter melting and open complex formation (reviewed in (Kaplan, 2013)). A directional model for TSS scanning is strongly suggested from mutational analysis of start site regions and the distribution of TSS when efficient start sites are compromised (Kostrewa et al., 2009; Kuehner and Brow, 2006). The contribution of scanning to TSS usage in other organisms is unknown; however, the majority of promoters in higher eukaryotes utilize multiple, dispersed TSS in a manner at least superficially analogous to *S. cerevisiae* (Choi et al., 2002; Consortium et al., 2014; Haberle et al., 2014; Hoskins et al., 2011).

GTFs TFIIB (encoded by *SUA7* in *S. cerevisiae*) and TFIIF (encoded by *TFG1, TFG2,* and *TFG3* in *S. cerevisiae*) contribute to TSS selection. TFIIB bridges the TATA binding protein (TBP)-promoter DNA complex and Pol II, and likely stabilizes open complex formation through interacting with single-stranded DNA sequences in the PIC; TFIIF guides and stabilizes Pol II binding during assembly of PIC, and appears to

functionally interact with TFIIB, and may directly regulate Pol II activity (Chen and Hampsey, 2004; Chen and Hahn, 2004; Eichner et al., 2010; Hampsey, 1998a; Henry et al., 1994; Luse, 2012; Sun and Hampsey, 1995) . A number of *tfg1* and *tfg2* alleles have been shown to shift distribution of TSS towards upstream positions (Eichner et al., 2010; Freire-Picos et al., 2005; Ghazy et al., 2004; Hahn and Young, 2011; Khaperskyy et al., 2008; Majovski et al., 2005). Conversely, mutations in *SUA7* generally have been shown to alter TSS distribution towards downstream positions (Chen and Hampsey, 2004; Faitar et al., 2001; Hull et al., 1995; Pinto et al., 1992; Pinto et al., 1994; Sun and Hampsey, 1995; Wu et al., 1999). Combination of TFIIB and TFIIF alleles can confer mutual suppression of their respective TSS defects along with TFIIF alleles' suppression of TFIIB alleles' temperature sensitive phenotypes (Freire-Picos et al., 2005; Ghazy et al., 2004; Sun and Hampsey, 1995). In addition, alleles of *SSL2*, which encodes an ATPase/helicase enzymatic subunit of TFIIH, have been shown to shift distribution of TSS towards upstream slightly, and one allele that was shown to partially suppress downstream TSS shifts the cold sensitivity of *sua7-1*, suggesting functional importance of Ssl2 in TSS selection (Goel et al., 2012). Alleles of some Pol II subunits have been shown to affect TSS usage distribution on their own and that of GTF alleles when combined. Combination of alleles in *tfg1* and *rpo21/rpb1* resulted both in suppressed temperature sensitivity and TSS defects of an *rpo21/rpb1* allele (FREIRE-PICOS *et al.* 2005). Combination of a *tfg1* allele and *rpb9*Δ resulted in exacerbated TSS defects and temperature sensitivity (Ghazy et al., 2004). Alleles of *rpb2* and *rpb9* were shown to suppress the downstream shift effect and cold temperature sensitivity of certain *sua7*

alleles (Chen and Hampsey, 2004; Sun and Hampsey, 1996; Sun et al., 1996). Thus, for some GTF alleles, effects on TSS selection parallel their effects on growth phenotypes when combined with *rpb* alleles.

Our characterization of the relationship between Pol II catalytic activity mutants and TSS defects suggested an activity-based framework for interpretation of Pol II mutant TSS defects (Braberg et al., 2013; Kaplan et al., 2012). A mechanistic explanation for the connection of Pol II activity and TSS selection will be critical for understanding Pol II initiation. Mutations in Pol II subunit-encoding genes *RPO21/RPB1*, *RPB2, RPB7* and *RPB9* have been previously shown to alter TSS utilization *in vivo*, however the mechanism of the alteration has been unclear (Braberg et al., 2013; Chen and Hampsey, 2004; Chen et al., 2007; Freire-Picos et al., 2005; Hull et al., 1995; Kaplan et al., 2012; Sun and Hampsey, 1996; Sun et al., 1996). In previous work, it was shown that *rpo21/rpb1* mutants with substitutions in the trigger loop (TL), a mobile portion of the Pol II active center, have altered elongation rates *in vitro.* One class of these Pol II mutants confers faster elongation rates (termed gain of funtion, GOF) *in vitro,* another class confers slower elongation rates (termed loss of function, LOF), and the two classes are generally mutually suppressive when combined. These Pol II catalytic activity mutants conferred various phenotypes both *in vitro* and *in vivo*, including altered TSS selection, RNA splicing, presumed termination or processing defects, and chromosome segregation defects (Braberg et al., 2013; Kaplan et al., 2012; Kaplan et al., 2008; Kireeva et al., 2008; Larson et al., 2012; Viktorovskaya et al., 2013).

The severity of TSS defects *in vivo* in both GOFs and LOFs correlated well with the extent of their deviation from WT elongation rate *in vitro*.

We found previously that Pol II GOF mutants shifted the distribution of TSS upstream at *ADH1* and other genes, similarly to *tfg2* alleles and *rpb9*Δ ; conversely, Pol II LOF mutants shifted distribution of TSS downstream at *ADH1*, similarly to most *sua7* alleles (Braberg et al., 2013; Kaplan et al., 2012). Directional alteration of *ADH1* TSS distribution by Pol II mutants, both GOFs and LOFs, mimic their effects on TSS distributions genome wide (H. Jin and C.D. Kaplan, unpublished results). Just as with the severity of their TSS defects, these Pol II mutants have growth defects *in vivo* that correlate with the extent of Pol II activity alteration. Mutants that have more severely altered activity *in vitro* (both fast and slow) show greater growth defects, more genetic intreractions, and greater alterations to gene expression profiles *in vivo*. Growth defects can be suppressed when Pol II LOF and GOF mutations are combined within the same enzyme; similarly, there is mutual suppression of Pol II mutant TSS distribution defects at *ADH1* in the double mutant, indicating a correlation between TSS defects and general growth defects (Kaplan et al., 2012). Since most Pol II mutant phenotypes we have studied correlate with strength of observed biochemical defects, it is difficult to distinguish whether observed *in vivo* growth defects derive especially from defects in a particular facet of transcription. Through genetic experiments and examination of TSS selection *in vivo*, we have attempted to understand further the relationship between Pol II activity defects, GTF function, and transcription initiation in *S. cerevisiae*. We also set out to extend our previous studies of factors that genetically interact with Pol II (genetic

interactors) (Braberg et al., 2013) to understand their roles in TSS selection and the relationship between initiation and growth defects of Pol II alleles.

Materials and methods

Yeast strains and media

Plasmids containing *tfg2Δ146-180*, *tfg2Δ261-273*, *tfg2Δ233-248* alleles were gifts from Steve Hahn (Eichner et al., 2010), plasmids containing *sua7-1*, *sua7-3* were gifts from Michael Hampsey (Chen and Hampsey, 2004). *sua7-58A5* and *sua7-70A5* alleles were generated by Quickchange site directed mutagenesis according to directions of the manufacturer (Stratagene/Agilent). Fragments containing target alleles were cloned into the yeast integrating vector pRS306 (Sikorski and Hieter, 1989) and transformed into a strain background used for phenotyping and primer extension assay. Yeast media used in phenotyping assays were made as previously described (Amberg et al., 2005; Braberg et al., 2013; Kaplan et al., 2012). YP media contained yeast extract (1% w/v, BD), peptone (2% w/v, BD), bacto agar (2% w/v, BD) supplemented with adenine and tryptophan. YPD media contained dextrose (2% w/v, VWR), YPRaf media contained raffinose (2% w/v, Amresco), YPRafGal media contained raffinose (2% w/v) and galactose (1% w/v, Amresco) as carbon source. YPRaf and YPRafGal media also contained antimycin A (1µg/ml, Sigma-Aldrich). Synthetic complete media were made with "Hopkins mix" with certain amino acids dropped out at the concentrations described in (Kaplan et al., 2012), after slight modifications of (Amberg et al., 2005).

SC-Leu+MPA media contains 20µg/ml final concentration of mycophenolic acid (Sigma-Aldrich) from a 10mg/ml concentrated stock in ethanol (stored at -20˚C). Mycophenolic acid (MPA) lowers cellular concentration of GTP by inhibition of IMPDH activity and induces expression of *IMD2*, which encodes an MPA-resistant form of *IMPDH*. Transcription mutants that are sensitive to lower GTP levels or those defective in induction of *IMD2* confer MPA sensitivity (MPA^S). For Pol II trigger loop mutants, MPA sensitivity is predictive of upstream start site defects at *ADH1* (Braberg et al., 2013).

Strains used in this study contain the *lys2-128∂* allele (Simchen et al., 1984) that renders cells auxotrophic for lysine due to a Ty1 retroelement long terminal repeat (LTR) insertion in the 5′ end of *LYS2.* Mutants that alter transcription at the allele can grow on medium lacking lysine (*e.g.* SC-Lys), a phenotype referred to as Spt^- (Suppressor of Ty). The *gal10Δ56* allele (Greger et al., 2000; Kaplan et al., 2005) comprises a deletion in the *GAL10* 3′-UTR, resulting in compromised RNA processing and termination at *GAL10*, allowing transcription readthrough downstream into *GAL7.* Lack of *GAL7* gene product allows accumulation of toxic intermediate products in galactose metabolism, thus WT *gal10Δ56* cells are sensitive to presence of galactose in the medium when *GAL* genes are expressed even in the presence of an additional usable carbon source (*e.g.* YPRafGal). Mutants that alter readthrough from *gal10Δ56* or otherwise increase *GAL7* expression show galactose resistance on YPRafGal media, a phenotype referred to as Gal^R.

Primer extension assay for start site utilization detection

Primer extension assays were performed as previously described (Ranish and Hahn, 1991) with a few modifications described in (Kaplan et al., 2012). Briefly, 30µg total RNA purified as previously described (Schmitt et al., 1990) was used to anneal with ^{32}P-labeld oligonucleotide priming downstream of *ADH1* start sites in 15µl total reaction volume. Reverse transcription reaction was performed by M-MLV reverse transcriptase (Fermentas) in the presence of RNase Inhibitor (Fermentas) in 45µl total reaction volume. Products were precipitated, digested with RNase A and separated in 8% acrylamide gel made with 19:1 acrylamide:bisacrylamide (Bio-Rad), 1XTBE and 7M urea, followed by visualization by phosphorimaging (Bio-Rad) and quantification with ImageQuant 5.1 software (GE).

Heatmaps for genetic interaction phenotypes

Growth on each media were scored using a 0-5 scoring system (0=no growth, 5=WT growth for all media except YPRafGal and SC-Lys; 0=WT growth, 5=growth of the mutant with maximum growth on the corresponding plate for YPRafGal and SC-Lys). To indicate growth difference in different growth conditions, all mutants on YPD, YPD 37°C, YPRaf, SC-Leu phenotypes were normalized to WT on each plate by subtraction (mutant score-WT score). Negative numbers (slower growth) are shown as blue, positive numbers (faster growth) are shown in red, inviable double mutants are colored in dark gray. Growth differences on SC-Leu+MPA were normalized to growth difference on SC-Leu by subtracting difference on SC-Leu from difference on SC-Leu+MPA thus rendering net growth difference due to MPA sensitivity/resistance.

Differences on YPRafGal and SC-Lys (WT growth is zero) were normalized to growth difference on YPD and SC-Leu ("standard growth condition" controls for these phenotyping media) by dividing difference on YPRafGal or SC-Lys by ratio of WT growth to mutant growth on YPD or SC-Leu to quantify resistance phenotypes. Mutants that have Gal^R or Spt^- phenotypes are thus shown as red in the heatmaps. Calculated score difference tables were turned into heat maps using GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/index.html).

Results

Allele-specific genetic interactions between GTF mutants and Pol II trigger mutants

We used changes in model gene *ADH1* TSS distribution as a proxy for *in vivo* initiation defects. To quantify changes in TSS distribution at *ADH1*, signals from *ADH1* TSS were placed into six bins and alterations in the fraction of TSS present in each bin were determined relative to the WT distribution (**Figure 3-1** A). We first examined how GTF mutants –known to alter TSS on their own– altered TSS defects of Pol II mutants and whether they modulated Pol II mutant growth phenotypes to explore their possible influence on TSS defects of Pol II mutants, and characterize any effects in light of any genetic interactions between Pol II mutants and GTF alleles. We wished to determine if opposite shifting Pol II and GTF TSS mutants were suppressive or additive when combined; for example, similarly to the combination of TFIIB and TFIIF alleles or combination of Pol II GOF and LOF mutants exhibiting suppression of TSS defects and growth phenotypes. Conversely, we might observe non-additive behavior in double

mutants, indicative of bypass or epistasis as we observed between *sub1*Δ and Pol II GOF alleles (Braberg et al., 2013). If a double mutant has a defect in growth phenotype that is better than expected from examination of individual phenotypes of single mutants (based on a multiplicative model for double mutant growth interactions (Schuldiner et al., 2006)) or an additive model for TSS defects, such an observation can be an example of epistasis. In such cases, single mutants would show a lack of independence when combined, with double mutants exhibiting phenotypes of one or the other single mutant, or a phenotype worse than either single mutant but to a lesser degree than would be expected from independently acting mutations. Finally, we wished to ask if GTF-Pol II genetic interactions and growth phenotypes strictly correlated with any observed modulation of Pol II mutant TSS defects.

We integrated three deletion mutants of *TFG2* into a yeast strain designed for phenotyping *rpo21/rpb1* alleles. *tfg2*Δ*146-180* and *tfg2*Δ*261-273* had been shown to shift *ADH1* TSS upstream, while *tfg2*Δ*233-248* had been shown to exhibit a mild *ADH1* TSS phenotype at best (Eichner et al., 2010). We also utilized *SUA7* alleles containing

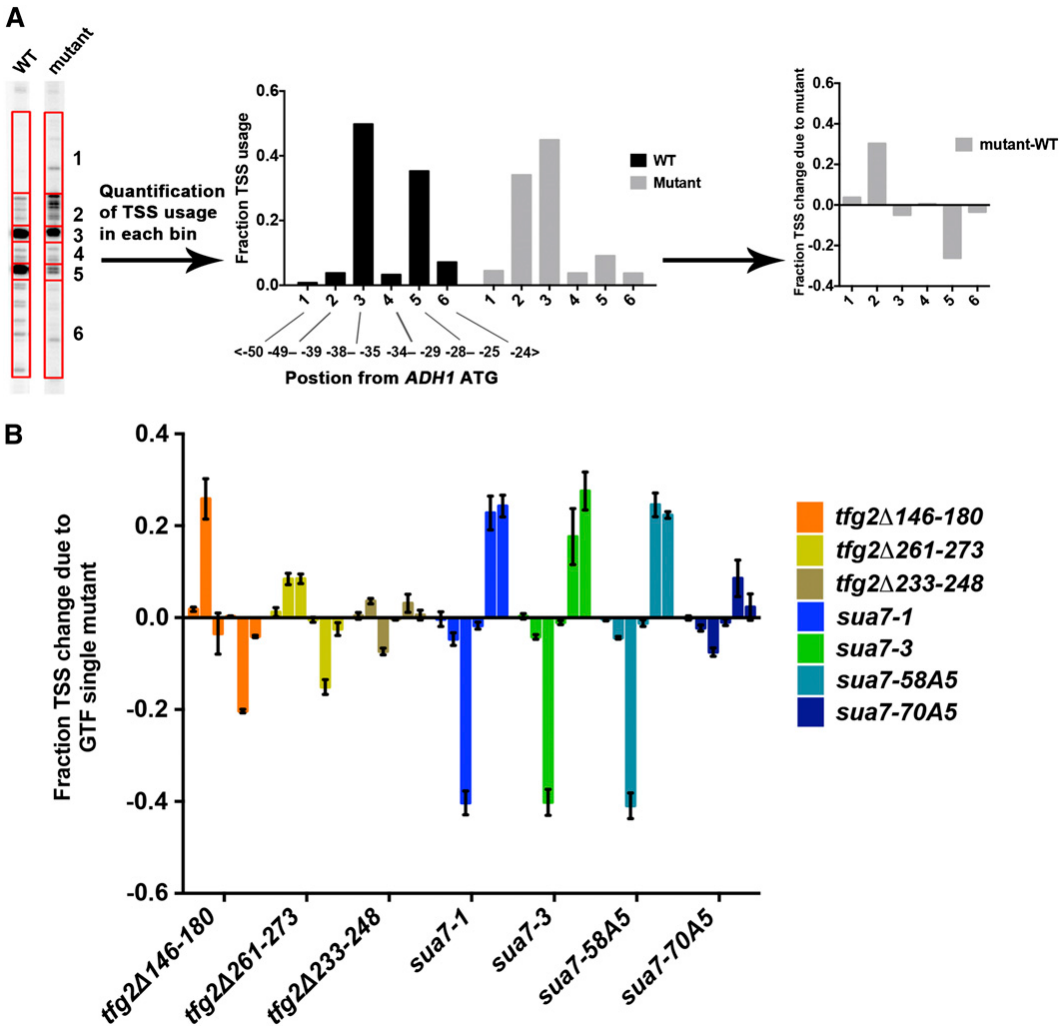


Figure 3-1. Transcription start site (TSS) usage distribution at *ADH1* and its alteration by Pol II GTF mutants.

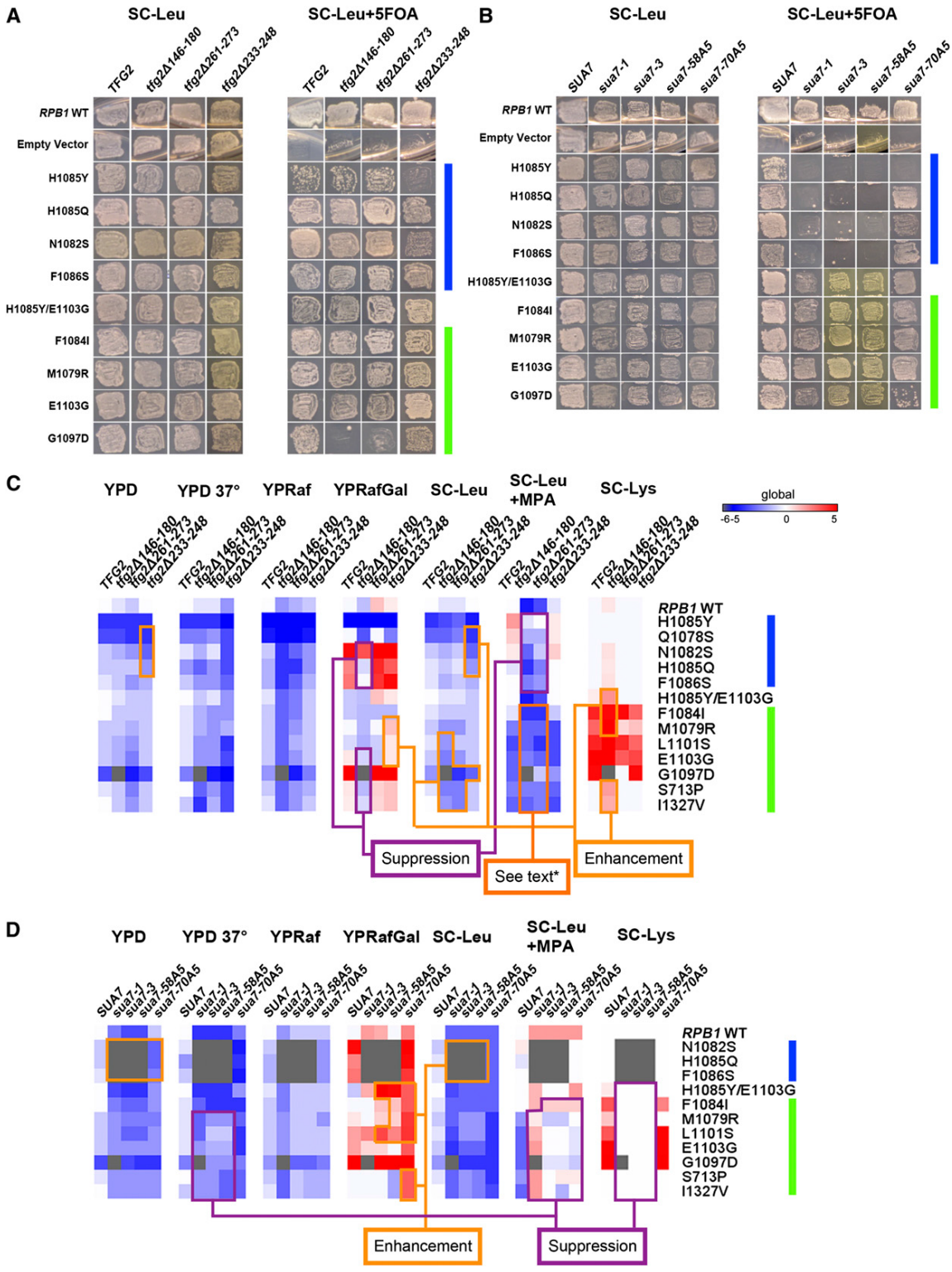
A. TSSs detected by primer extension at *ADH1* are distributed over a range of positions. To quantify TSS distributions, *ADH1* start site signals were divided into six bins separated by promoter position and normalized to total signal for each lane (left panel). TSS usage distributions were quantified for different strains (middle panel). Relative change in normalized TSS usage distribution for mutant compared to WT (negative numbers indicate relative decrease in TSS position usage, positive numbers indicate relative increase) is then calculated and plotted (right panel). **B.** Alterations in TSS usage at *ADH1* caused by each GTF mutant shown were quantified as in A. Start site defects of these GTF mutants are consistent with previous publications (Eichner et al., 2010; Wu et al., 1999), except for *sua7-A5* alleles, which are in contrast to (Zhang et al., 2002). Graphs show average of at least three independent determinations with error bars representing standard deviations. See Figure 3-6 for representative primer extension experiments.

substitutions in the "B-reader region": *sua7-1* (encodes E62K in TFIIB), *sua7-3* (encodes R78C in TFIIB) alleles that have been shown to confer downstream TSS shifts by the Hampsey group (Pinto et al., 1994; Wu et al., 1999), and *sua7-58A5* and *sua7-70A5* mutants previously described as having upstream shift effects (Zhang et al., 2002). *sua7-58A5* and *sua7-70A5* each contain an insertion of five alanines at different positions in the B-reader (amino acid 58 or 70, respectively) and were recreated in our lab based on the published description in Zhang *et al*. (Zhang et al., 2002). Effects of GTF single mutants on *ADH1* TSS distribution are shown in **Figure 3-1** B. *tfg2Δ146-180* has stronger upstream shifts at *ADH1* than *tfg2Δ261-273*, while *tfg2Δ233-248* has little effect, consistent with (Eichner et al., 2010). *sua7-1* and *sua7-3* show strong downstream shifts as shown previously (Pinto et al., 1994; Wu et al., 1999), however, *sua7-58A5* and *sua7-70A5* both show downstream shifts, strong and weak respectively, which is typical behavior of *sua7* TSS mutants but is opposite of published observations. We cannot explain this discrepancy but consider *sua7-58A5* and *sua7-70A5* to have standard behavior for *sua7* alleles.

We tested the growth of these GTF mutants alone or in combination with Pol II mutants by transformation and plasmid shuffling of *rpo21/rpb1* alleles in place of *RPO21*. Genetic interactions between GTF alleles and Pol II mutants appear complex, and we will first describe genetic interactions apparent on growth of strains in rich or defined media (**Figure 3-2** A-B, YPD and SC-Leu media in C-D), followed by observed genetic interactions relating to gene-specific transcriptional phenotypes (Spt^-, Gal^R, MPA sensitivity) (**Figure 3-2** C-D).

When GTF alleles and Pol II mutants were combined, we observed allele-specific interactions between GTF alleles and Pol II GOF/LOF mutant classes. When Pol II GOF mutants and *tfg2* alleles were combined, the most severe GOF mutant *rpo21/rpb1* G1097D, exhibited a strong negative interaction with *tfg2*Δ*146-180* or *tfg2*Δ*261-273* (lethality and synthetic sickness, respectively). LOF Pol II mutants (downstream TSS shifting mutants) and *tfg2*Δ*146-180* and *tfg2*Δ*261-273* (upstream TSS shifting mutants) did not result in suppression of growth defects of Pol II LOF mutants on standard rich or defined media, in contrast to mutual suppression of growth defects when Pol II GOF and LOF (opposite TSS shifting mutants) are combined (Kaplan et al., 2012), or when *tfg* and *sua7* allleles (opposite TSS shifting mutants) have been combined (Freire-Picos et al., 2005; Ghazy et al., 2004; Sun and Hampsey, 1995) (**Figure 3-2** A, C). The *tfg2*Δ*233-248* allele, which does not have a clear TSS defect, exhibited a negative interaction with Pol II LOF alleles but no clear interactions with Pol II GOF alleles. Neither *tfg2* upstream shifting allele could rescue lethal LOF Pol II mutants (**Figure 3-3** A), in contrast to rescue of lethal Pol II LOFs when combined with GOF mutants (Kaplan et al., 2012).

Figure 3-2. Genetic interactions between GTF and Pol II mutants.
A. Plasmids containing *tfg2* mutants were integrated into yeast strains constructed to allow shuffling of a WT *RPO21/RPB1 URA3* plasmid in favor of WT or mutant *rpo21*/*rpb1 LEU2* plasmids through use of 5FOA poisoning of $URA3^{+}$ cells. On the left, SC-Leu media allows coexistence of both *URA3* and *LEU2* plasmids while on the right, supplementation of SC-Leu with 5FOA uncovers *rpb1* phenotypes by selecting against the *RPB1 URA3* plasmid. Pol II mutants that have slower elongation rate than WT *in vitro* (LOFs) and mutants that genetically cluster with them are annotated with a blue bar, mutants that have faster elongation rate than WT *in vitro* (GOFs) and mutants genetically cluster with them are labeled with a green bar. Mutants are arranged by their measured elongation rates or elongation rates inferred by strength of genetic phenotypes compared to those mutants tested biochemically (Kaplan et al., 2012). **B.** *sua7* allele-Pol II mutant interactions examined as for *tfg2* alleles in A. C-D. Phenotypes of viable GTF-Pol II double mutants are shown as a heat map with qualitative determinations of growth defects on various media. Inviable double mutants are colored in gray. In YPD, YPD 37˚C, YPRaf and SC-Leu media, single and double mutant growth levels are normalized to WT. Blue indicates decreased growth relative to WT, red indicates increased growth compared to WT. In SC-Leu+MPA, growth difference is normalized to that on SC-Leu to quantify MPA sensitivity (shown as blue) or resistance (shown as red). In YPRafGal and SC-Lys, Growth on the plate is divided by ratio of WT growth to mutant growth on corresponding general media (YPD and SC-Leu) to account for Gal^{R} and Spt^{-} phenotypes (Gal^{+} and Lys^{+}, shown in red) in contrast to their underlying growth defects. **C.** *tfg2* mutants **D.** *sua7* mutants. See Figure 3-3, 3-4 for representative spot growth assay figures used for quantification in heatmaps and Materials and Methods for further explanation of heatmaps.


A
SC-Leu
SC-Leu+5FOA
TFG2
tfg2Δ146-180
tfg2Δ261-273
tfg2Δ233-248
RPB1 WT
Empty Vector
H1085Y
H1085Q
N1082S
F1086S
H1085Y/E1103G
F1084I
M1079R
E1103G
G1097D
B
SC-Leu
SC-Leu+5FOA
SUA7
sua7-1
sua7-3
sua7-58A5
sua7-70A5
RPB1 WT
Empty Vector
H1085Y
H1085Q
N1082S
F1086S
H1085Y/E1103G
F1084I
M1079R
E1103G
G1097D
C
YPD
YPD 37°
YPRaf
YPRafGal
SC-Leu
SC-Leu +MPA
SC-Lys
global
-6-5
0
5
RPB1 WT
H1085Y
Q1078S
N1082S
H1085Q
F1086S
H1085Y/E1103G
F1084I
M1079R
L1101S
E1103G
G1097D
S713P
I1327V
Suppression
See text*
Enhancement
D
YPD
YPD 37°
YPRaf
YPRafGal
SC-Leu
SC-Leu +MPA
SC-Lys
RPB1 WT
N1082S
H1085Q
F1086S
H1085Y/E1103G
F1084I
M1079R
L1101S
E1103G
G1097D
S713P
I1327V
Enhancement
Suppression

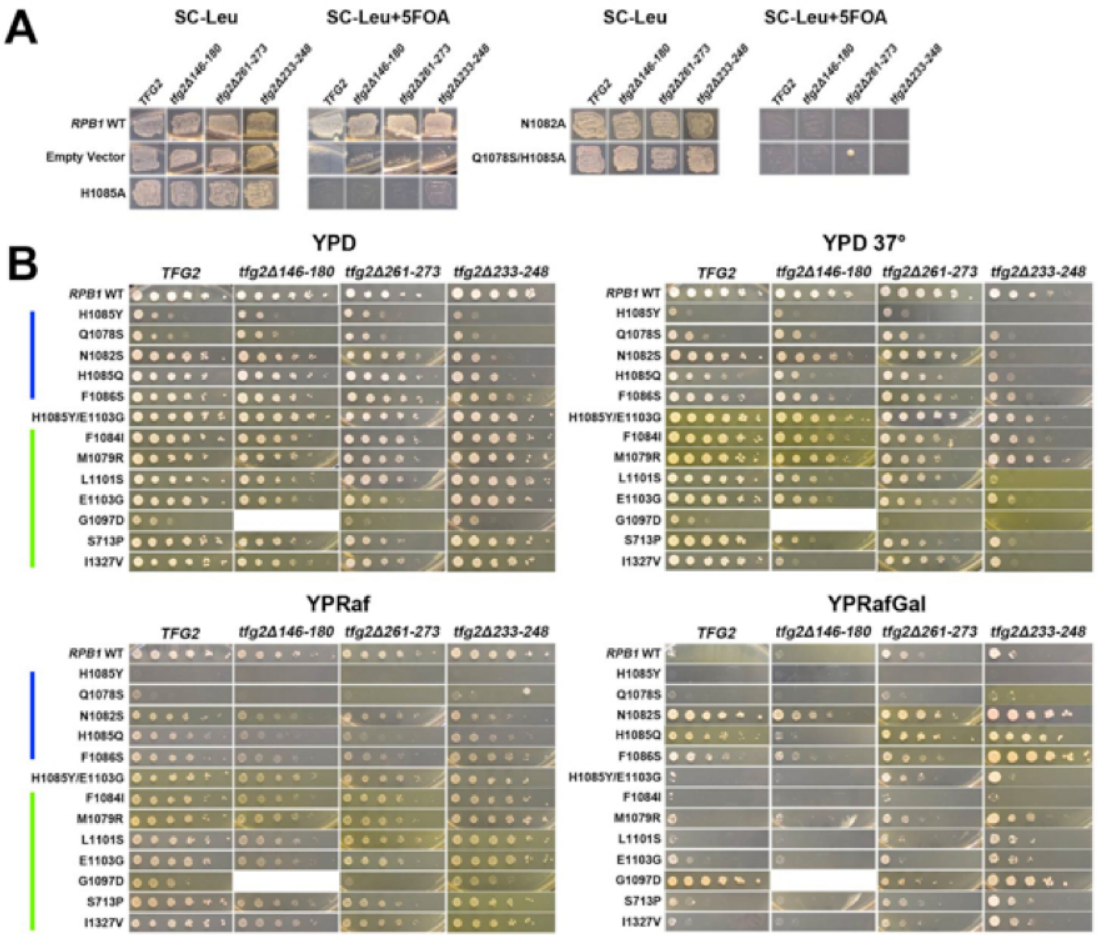


Figure 3-3. Phenotypes of *tfg2* alleles in combination with Pol II alleles.
A. Inability of *tfg2* alleles to rescue lethal LOF Pol II alleles detected by growth of double mutants on 5FOA (see Methods and Materials for details). **B.** Serial dilutions of viable *tfg2*/*rpo21* (*rpb1*) double mutant alleles on various media for phenotyping of genetic interactions (general growth, temperature sensitivity, MPA^S, Gal^R and Spt^- phenotypes). LOF Pol II alleles are marked by blue bar, GOF by green. Heatmap presentation of phenotype quantifications of this assay is shown in Figure 3-2 C.

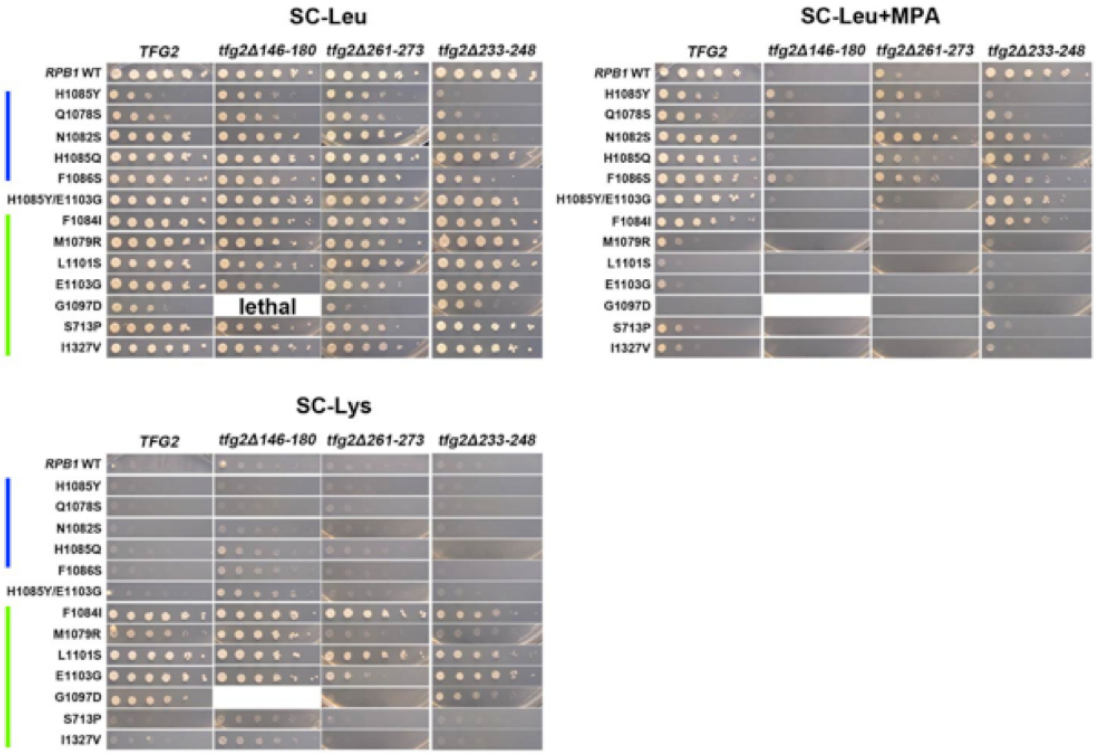


Figure 3-3 Continued.

When *sua7* mutants were combined with GOF Pol II mutants (upstream TSS shifting mutants), we observed apparent lack of additivity in growth defects of strong downstream shifting *sua7* alleles combined with Pol II GOFs. The double mutants generally showed growth defects in between those of the single mutants, suggesting directional suppression or epistasis, but not mutual suppression, where double mutants would be expected to grow better than either single mutant (mutual suppression is generally observed when Pol II LOF and GOF trigger loop mutants combined within the same enzyme). These effects were more pronounced on YPD medium than on YPRaf or defined medium (SC-Leu). We also observed partial suppression of the sensitivity to increased temperature (37˚C, Ts^- phenotype) of *sua7-1, sua7-3* and *sua7-58A5* alleles by Pol II GOF alleles (**Figure 3-2**, **Figure 3-4**). The weak downstream shifting allele *sua7-70A5* enhanced growth defects with Pol II GOF alleles, distinct from the other stronger *sua7* alleles tested.

In contrast to the milder phenotypes of *sua7*-Pol II GOF strains, Pol II LOF mutants were exquisitely sensitive to defects in TFIIB, as widespread synthetic lethality or sickness was observed between *sua7* alleles and Pol II LOF mutants. *sua7-1, sua7-3*, and *sua7-58A5* showed very strong negative interactions with all LOF Pol II mutants tested with most double mutants being inviable. A weak downstream shifting allele, *sua7-70A5*, showed negative but weaker interactions with Pol II LOF alleles. Both classes of mutant, Pol II LOF and *sua7* alleles alter TSS distributions in similar fashion, suggesting exacerbated TSS defects might underlie observed synthetic genetic interactions (see below). These results indicate that combinations of Pol II and GTF

mutants exhibiting the same polarity of TSS defects can lead to exacerbation of growth defects, wherein aggravated initiation defects may be a major contributor of the observed growth defects. The strongest genetic interaction between Pol II mutants and GTF alleles was observed when Pol II LOF mutants and strong *sua7* downstream shifting alleles were combined (lethality), suggesting that Pol II LOF mutants are much more sensitive to initiation defects than GOF mutants. Combinations of mutants with opposing TSS distribution defects resulted in partial but not necessarily mutual suppression of single mutant growth defects on generic media.

We also examined conditional growth phenotypes on several other types of media, including those reporting on gene-specific transcription defects *in vivo* (Braberg et al., 2013; Kaplan et al., 2012). Results are shown as a heatmap of normalized estimates of phenotypic strength as determined by visual determination of growth on plates (*tfg2* upstream shifting alleles in **Figure 3-2** C, *sua7* downstream shifting alleles in **Figure 3-2** D, see **Figure 3-3**, **Figure 3-4** for representative images, see Materials and Methods for calculations).

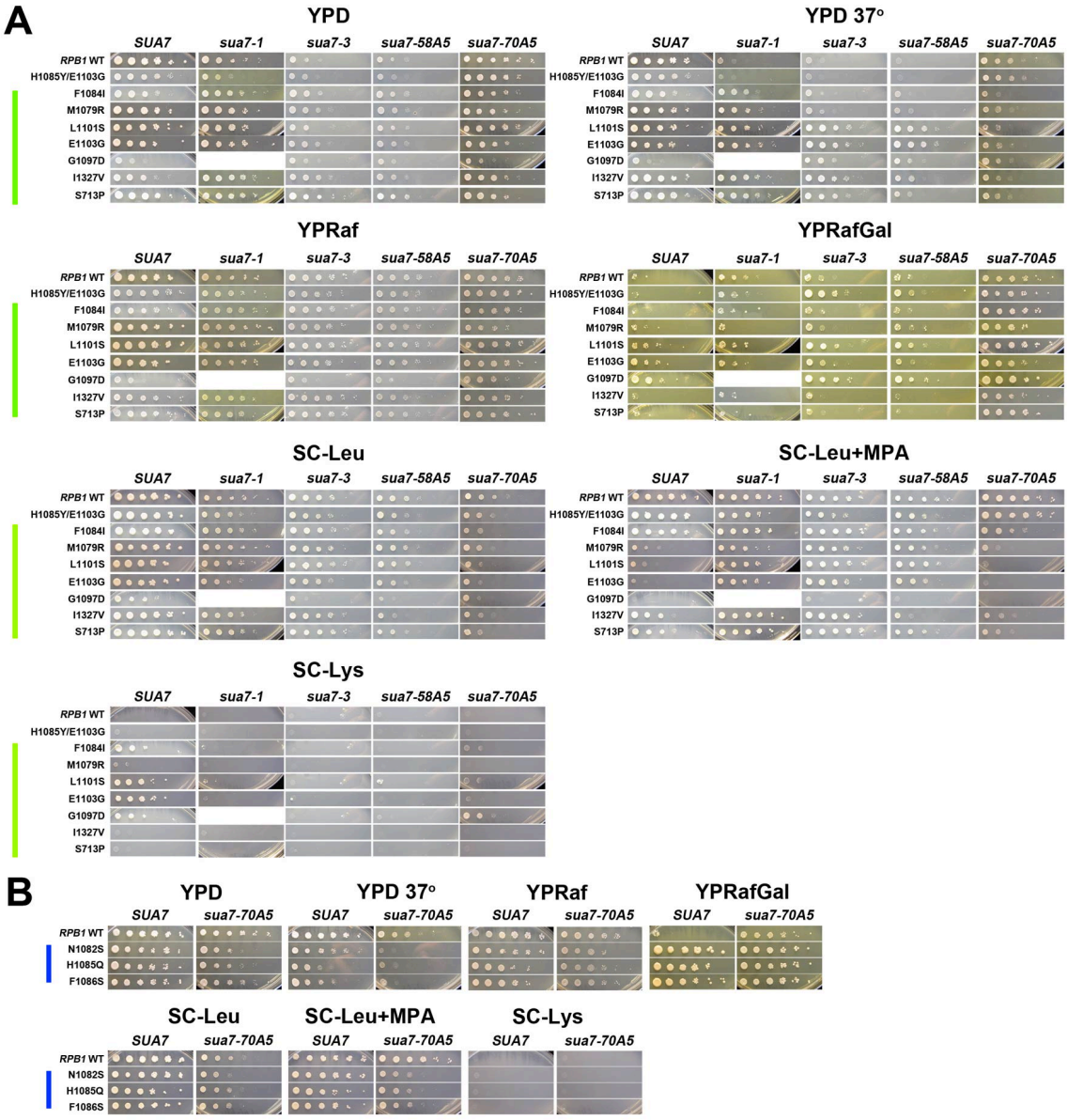


Figure 3-4. Phenotypes of *sua7* alleles in combination with Pol II alleles.
A. Serial dilutions of viable *sua7*/*rpo21* (*rpb1*) double mutant alleles on various media for phenotyping of genetic interactions (general growth, temperature sensitivity, MPA^S, Gal^R and Spt^- phenotypes). LOF Pol II alleles are marked by blue bar, GOF by green. Heatmap presentation of phenotype quantifications of this assay is shown in Figure 2D.
B. Spot growth assay for combinations of LOF Pol II alleles with *sua7-70A5*.

We observed sensitivity to mycophenolic acid (MPA) for *tfg2* alleles, likely corresponding to upstream TSS shifts causing inability to induce *IMD2* in the presence of MPA, with *tfg2*Δ*146-180* the most MPA sensitive. MPA sensitivity has been shown to correlate well with upstream TSS shifts of a subset of Pol II mutants, including those tested here (Braberg et al., 2013). Conversely, *sua7* alleles appeared resistant to MPA, similar to Pol II LOF alleles (**Figure 3-2** C-D). When Pol II GOF alleles are combined with the strong upstream shifting allele *tfg2*Δ*146-180,* the double mutants are inviable on this medium. The weaker upstream shifting allele *tfg2*Δ*261-273* also exacerbates MPA sensitivity of Pol II GOF alleles, with the resulting double mutants exhibiting no growth on this medium. This enhancement of MPA sensitivity leading to double mutants' lack of detectable growth is not well illustrated in our heatmap due to our calculation metric being unable to capture zero growth during calculation of "net" MPA sensitivity (See Matrials and Methods). In contrast, MPA sensitivities of *tfg2* alleles were suppressed when combined with Pol II LOFs; similarly, MPA sensitivities of Pol II GOFs were suppressed when combined with all *sua7* alleles, being more apparent in stronger *sua7* downstream shifting alleles. Therefore, MPA phenotypes of combinations of GTF alleles and Pol II mutants appeared additive or suppressive depending on the nature of the allele class: when upstream shifing alleles were combined, MPA sensitivity was exacerbated; when an upstream shifting mutant was combined with a downstream shifting mutant, MPA sensitivity was alleviated (**Figure 3-2** C-D). These results support MPA sensitivity as a readout for initiation defects and not necessarily elongation defects, as widely assumed, and predicts that TSS defects of Pol II mutants and GTF alleles may be

additive and suppressive. *sua7* alleles with strong TSS defects showed strong temperature sensitivity (Ts^-), however this could partially be alleviated when combined with Pol II GOF mutants (**Figure 3-2** D).

We also examined gene-specific transcription related Gal^R and Spt^- phenotypes, which have less clear relationships to TSS defects (see Materials and Methods). Most Pol II GOF mutants do show Spt^- phenotype as measured by suppression of lysine auxotrophy (Lys^-) in the presence of *lys2-128∂* allele (see Materials and Methods), but there is only partial correlation with upstream shifting TSS defects. *tfg2Δ146-180* enhanced Spt^- phenotypes of all moderate Pol II alleles, while the other *tfg2* alleles did not show apparent interaction with Pol II alleles for the Spt^- phenotype. Stronger downstream shifting *sua7* alleles suppressed Spt^- phenotypes of Pol II GOF alleles (**Figure 3-2** C-D). A number Pol II LOF and GOF mutants have been shown to exhibit the Gal^R phenotype in the presence of the *gal10Δ56* allele of *GAL10* (see Materials and Methods for description)(Kaplan et al., 2012). The strongest upstream shifting allele *tfg2Δ146-180* suppressed Gal^R phenotypes of all Pol II alleles that had the Gal^R phenotype, whether GOFs or LOFs; *tfg2Δ233-248*, an allele with no apparent *ADH1* TSS defect, enhanced Gal^R phenotypes of weak Pol II GOF alleles. *sua7* alleles showed Gal^R phenotypes on their own, and enhanced those of Pol II GOF alleles (**Figure 3-2** C-D). The wide range of genetic interactions including enhancement and suppression of these conditional growth phenotypes suggests a complex network between Pol II and TFIIB/TFIIF that may relate to gene specific effects not apparent in overall double mutant growth phenotypes.

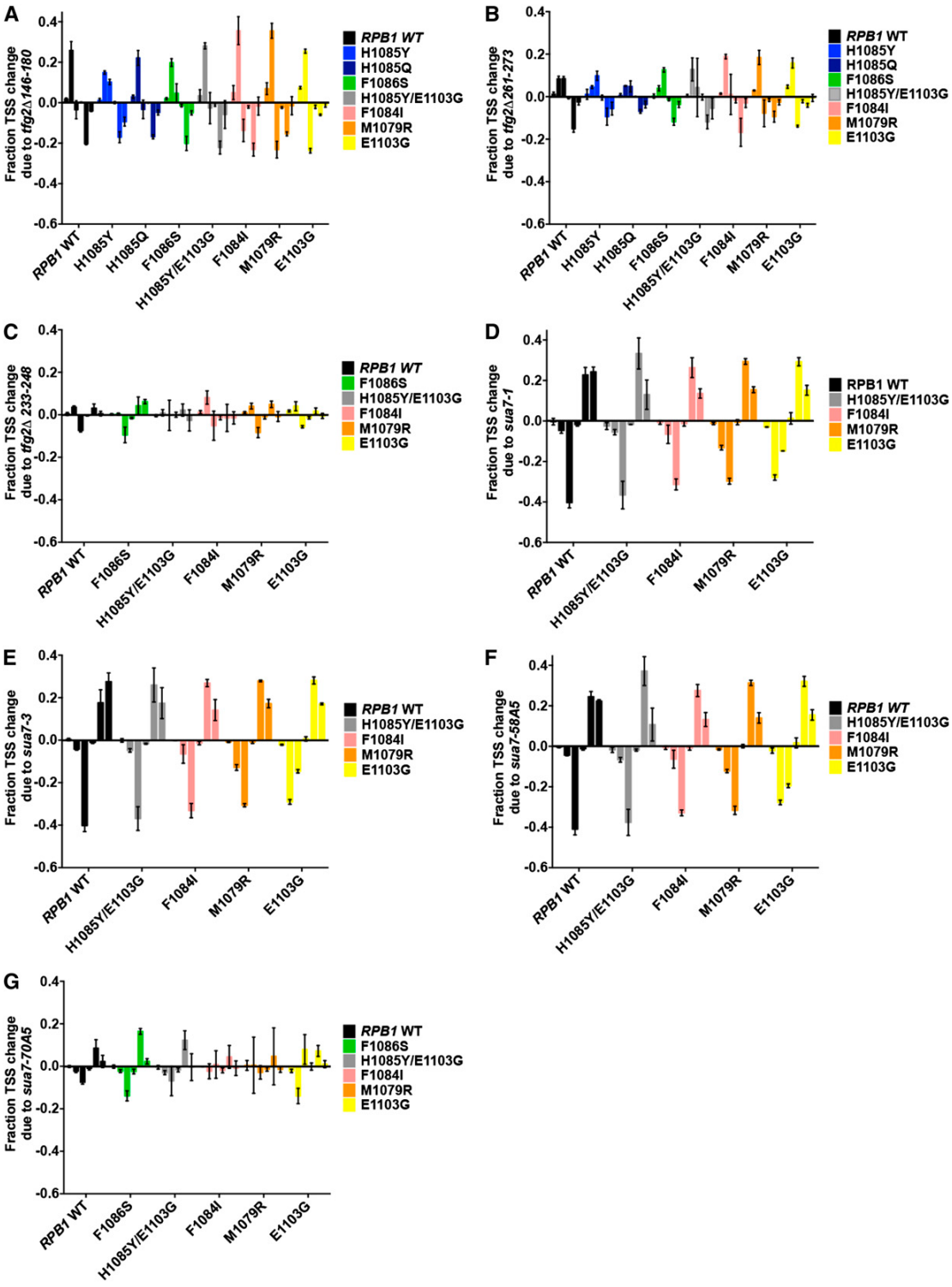
Combinations of GTF alleles and Pol II alleles lead to mutual suppression of TSS defects but not mutual suppression of generic growth defects

Because we observed above that GTF TSS defective mutants and Pol II TSS defective mutants conferred enhanced growth defects when same direction TSS shifting mutants were combined, but showed mostly weak, conditional, or directional suppression when opposite direction shifting mutants were combined, we examined how each class of double mutant affected TSS distribution at *ADH1*. This was done to determine if TSS defects of combination of GTF alleles and Pol II mutants were additive and suppressive similar to MPA-sensitivity phenotypes, or exacerbating but not mutually suppressive similar to general growth phenotypes of double mutants.

We observed that the TSS defects of GTF mutants and Pol II mutants were uniformly additive or suppressive, depending on the direction of TSS shifts of individual mutants, but not epistatic (**Figure 3-5**, see **Figure 3-6** for representative raw data). *tfg2*Δ*146-180* and *tfg2*Δ*261-273* mutants shifted *ADH1* TSS upstream relative to all single Pol II alleles tested, indicating exacerbation of Pol II GOF alleles that shift *ADH1* TSS upstream on their own, and suppression of Pol II LOF alleles that shift TSS downstream on their own (**Figure 3-5** A-B). The mutual suppression of TSS defects observed was similar to those observed for double mutant combinations of TFIIF and TFIIB alleles or for intra-Pol II double mutants. Conversely, those cases were accompanied by mutual suppression of growth phenotypes, which appears lacking for combinations of GTF alleles with Pol II trigger loop alleles (**Figure 3-2**). Additionally, strong downstream shifting alleles *sua7-1, sua7-3, sua7-58A5* shifted *ADH1* TSS of

sua7-Pol II GOF double mutants downstream relative to all Pol II GOF single mutants (**Figure 3-5** D-F), indicating additive effects of opposite polarity shifts, and therefore suppression of GOF Pol II allele TSS defects at *ADH1*. *sua7-70A5*, a weak downstream shifting allele, shifted *ADH1* TSS of *sua7-70A5*-Pol II double mutants downstream relative to all Pol II allele backgrounds (**Figure 3-5** G), indicating exacerbated downstream TSS shifts of LOF Pol II alleles and suppression of upstream TSS shifts of GOF Pol II alleles. TSS defects, generic and conditional growth phenotypes tested in GTF-Pol II double mutants suggest that TSS defects may contribute to general growth defects when TSS defects are severe, yet suppression of TSS defects (as measured at *ADH1*) does not correlate with suppression of Pol II mutant generic growth defects. Taken together, these results suggest that TSS defects may contribute to Pol II mutant growth defects, and Pol II activity alterations can partially compensate for defects in GTFs, but initiation defects are not likely to be the main or only drivers of observed Pol II allele growth phenotypes.

Figure 3-5. Modulation of Pol II mutant TSS selection defects at *ADH1* by GTF mutants.
A. Quantification of effects of GTF alleles on Pol II alleles on TSS utilization by comparison of double mutants to respective Pol II single mutants at *ADH1* (quantified as in **Figure 2-1** A). Values indicate average of a minimum of three independent determinations, with standard deviations represented by error bars. **A.** *tfg2Δ146-180.* **B.** *tfg2Δ261-273.* **C.** *tfg2Δ233-248* **D.** *sua7-1.* **E.** *sua7-3.* **F.** *sua7-58A5.* **G.** *sua7-70A5.* See Figure 3-6 for representative primer extension experiments.


A
Fraction TSS change due to tfg2Δ146-180
RPB1 WT
H1085Y
H1085Q
F1086S
H1085Y/E1103G
F1084I
M1079R
E1103G
B
Fraction TSS change due to tfg2Δ261-273
C
Fraction TSS change due to tfg2Δ 233-248
D
Fraction TSS change due to sua7-1
E
Fraction TSS change due to sua7-3
F
Fraction TSS change due to sua7-58A5
G
Fraction TSS change due to sua7-70A5

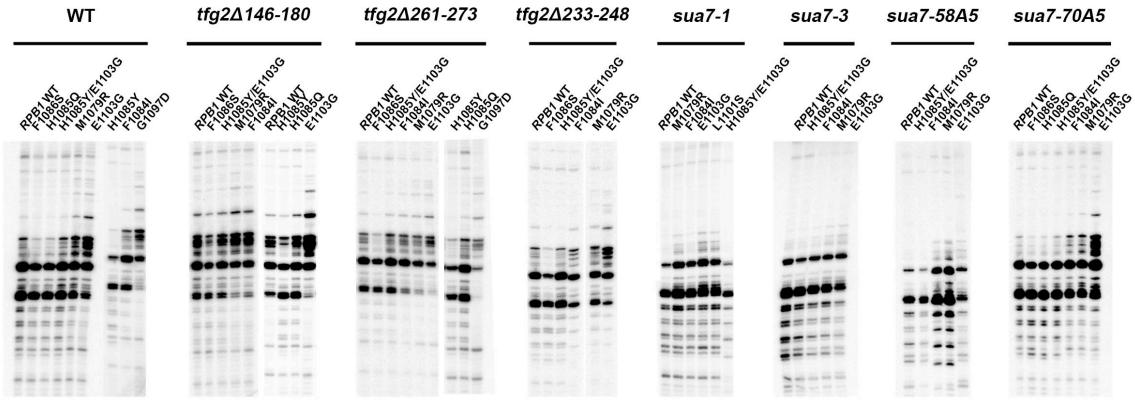


Figure 3-6. Pol II/GTF double mutant effects on *ADH1* transcription start site selection.

TSSs at *ADH1* detected by primer extension for various transcription mutant strains. One representative experiment of at least three independent replicates is shown. GTF allele description is shown on top with a bar that indicates lanes showing Pol II alleles combined with a particular GTF allele. Relevant mutation in Pol II is labeled above each lane. Quantification of these experiments is shown in Figure 3-5.

Genetic interactors with widespread genetic interactions with Pol II TSS defective alleles do not generally have TSS defects on their own or when combined with Pol II alleles

Inspired by the discovery of a Pol II genetic interactor, *sub1*Δ, that conferred a downstream TSS defect on its own, enhanced downstream TSS defects and growth defects of LOF Pol II alleles, yet showed epistasis with GOF Pol II alleles for TSS defects and growth phenotypes (Braberg et al., 2013), we investigated how factors that have genetic interactions with Pol II TSS defective alleles affected TSS distributions on their own or in combination with Pol II alleles. Are genetic interactions of factors and Pol II TSS shifting alleles predictive of their effects on TSS utilization? The genetic interactions of *dst1*Δ*, rtf1*Δ*, sgf73*Δ*, paf1*Δ*, and ctr9*Δ with Pol II alleles have been shown previously (Braberg et al., 2013; Hartzog et al., 1998; Malagon et al., 2006). These and additional factors are illustrated in **Figure 3-7**. Deletion of *DST1* (encoding the general transcription elongation factor TFIIS) showed allele-specific genetic interactions with GOF Pol II alleles and suppressed the Spt$^-$ phenotypes of Pol II alleles. *RTF1, CTR9, PAF1* (genes encoding subunits of the Paf1C complex) showed stronger genetic interactions with Pol II LOF alleles, exhibited mild Spt$^-$ phenotypes on their own, yet suppressed the Spt$^-$ phenotypes of Pol II alleles. On the other hand, *SGF73* (encoding a subunit of the histone-modifying SAGA complex) showed genetic interactions with both classes of alleles, enhanced MPA resistance of Pol II LOF alleles, and suppressed Spt$^-$ phenotypes of Pol II alleles (**Figure 3-7**). We found that deletions of these genetic interactors did not confer any strong TSS defects at *ADH1* on their own (**Figure 3-8** A), nor did they modulate TSS defects of either LOF or GOF Pol II alleles (**Figure 3-8** B).

The genetic interactions these factors exhibit with Pol II TSS-defective alleles may go through distinct mechanisms from *sub1*Δ, whose genetic interactions on growth mirrored its effects on TSS defects at *ADH1* when combined with Pol II alleles (Braberg et al., 2013). These results suggest genetic interactions between factors and Pol II alleles are not predictive of initiation defects, and that growth defects of many or most double mutant combinations with Pol II alleles do not result from exacerbation of TSS defects, under the assumption that *ADH1* is a proxy for global TSS defects (supported by our unpublished global analysis of TSS defects in Pol II mutants).

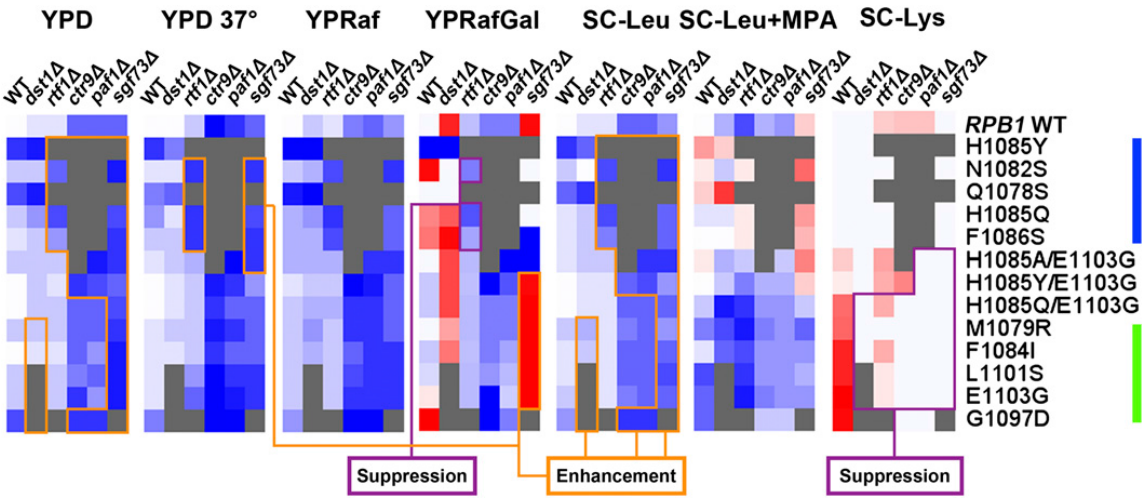


Figure 3-7. Genetic interactions between Pol II alleles and genetic interactor deletions.
Phenotypes of genetic interactor deletions on their own or in combination with Pol II mutants on different medium normalized to WT are shown in heatmap. See description in Fig 3-2 C-D and Materials and Methods for heatmap details. See Figure 2-30 and Figure 3-9 for representative spot growth assays used for heatmaps.

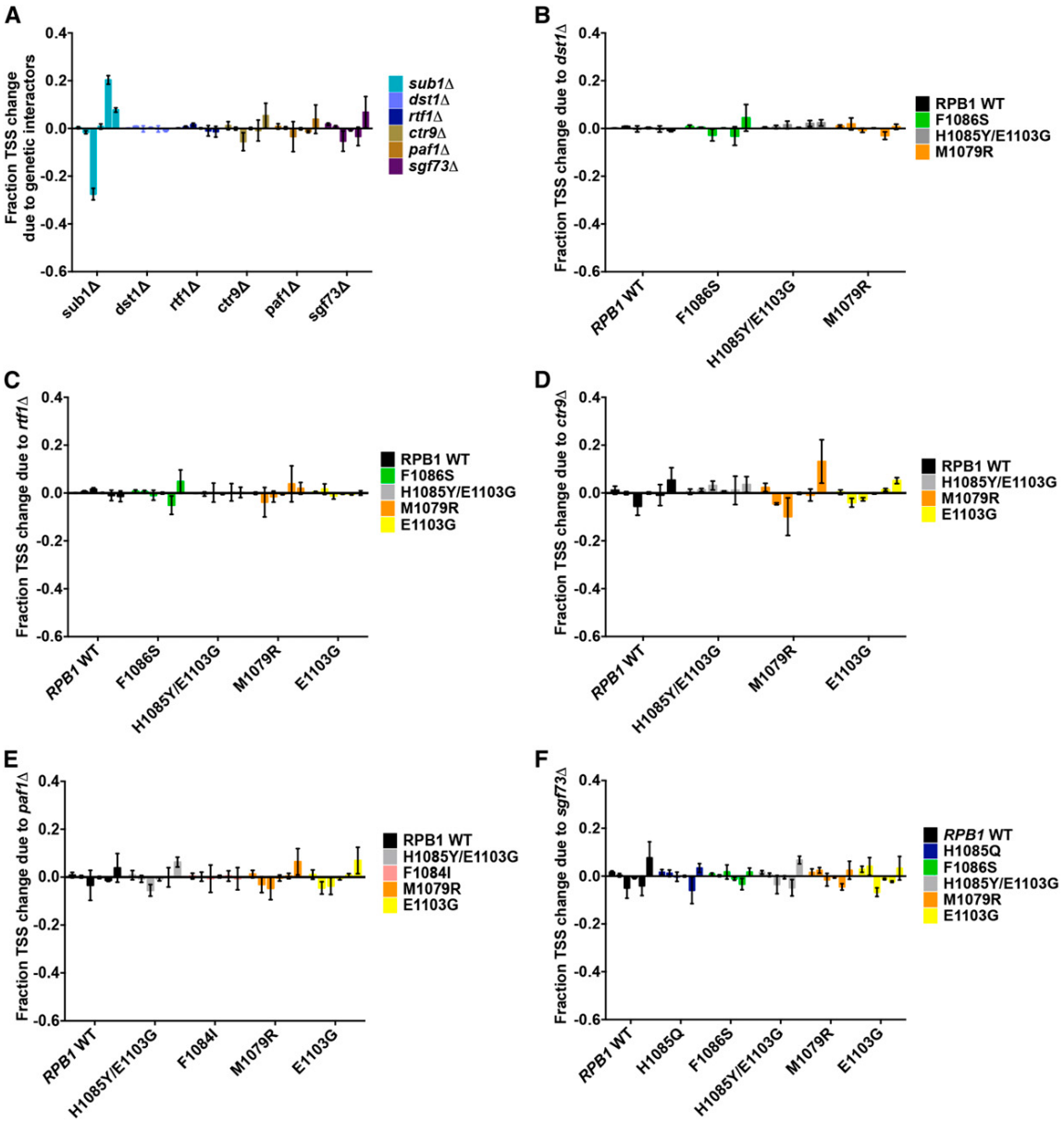


Figure 3-8. Genetic interactors do not generally modulate TSS defects of Pol II mutants at *ADH1*.

A. Quantification of TSSs usage alterations relative to WT at *ADH1* in genetic interactor deletions are shown (quantified as in Figure 1A). Values indicate a minimum of three independent determinations, with standard deviations represented by error bars. B-F. Quantification of TSSs usage distribution differences between genetic interactor deletion- Pol II double mutants relative to Pol II single mutants are shown. **B.** *dst1*Δ **C.** *rtf1*Δ. **D.** *ctr9*Δ. **E.** *paf1*Δ. **F.** *sgf73*Δ respectively. See **Figure 3-10** for representative primer extension experiments.

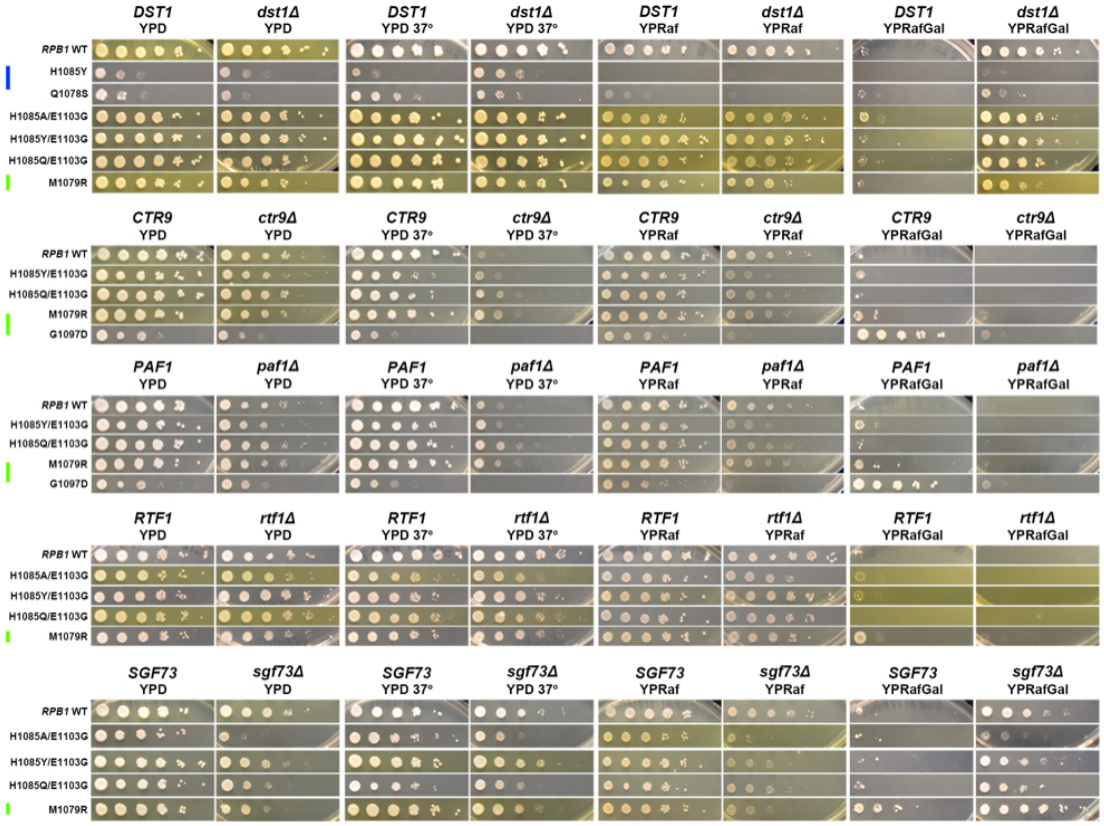


Figure 3-9. Serial dilutions of Pol II genetic interactor deletions combined with *rpo21 (rpb1)* alleles to examine genetic interactions on general growth and on transcription-related gene-specific phenotypes (Spt$^-$, MPA^S, Gal^R phenotypes). LOF Pol II alleles are marked by blue bar, GOF by green. Heatmap presentation of phenotype (general growth, temperature sensitivity, MPA^S, Gal^R and Spt$^-$ phenotypes) quantification of this assay is shown in Figure 3-7.

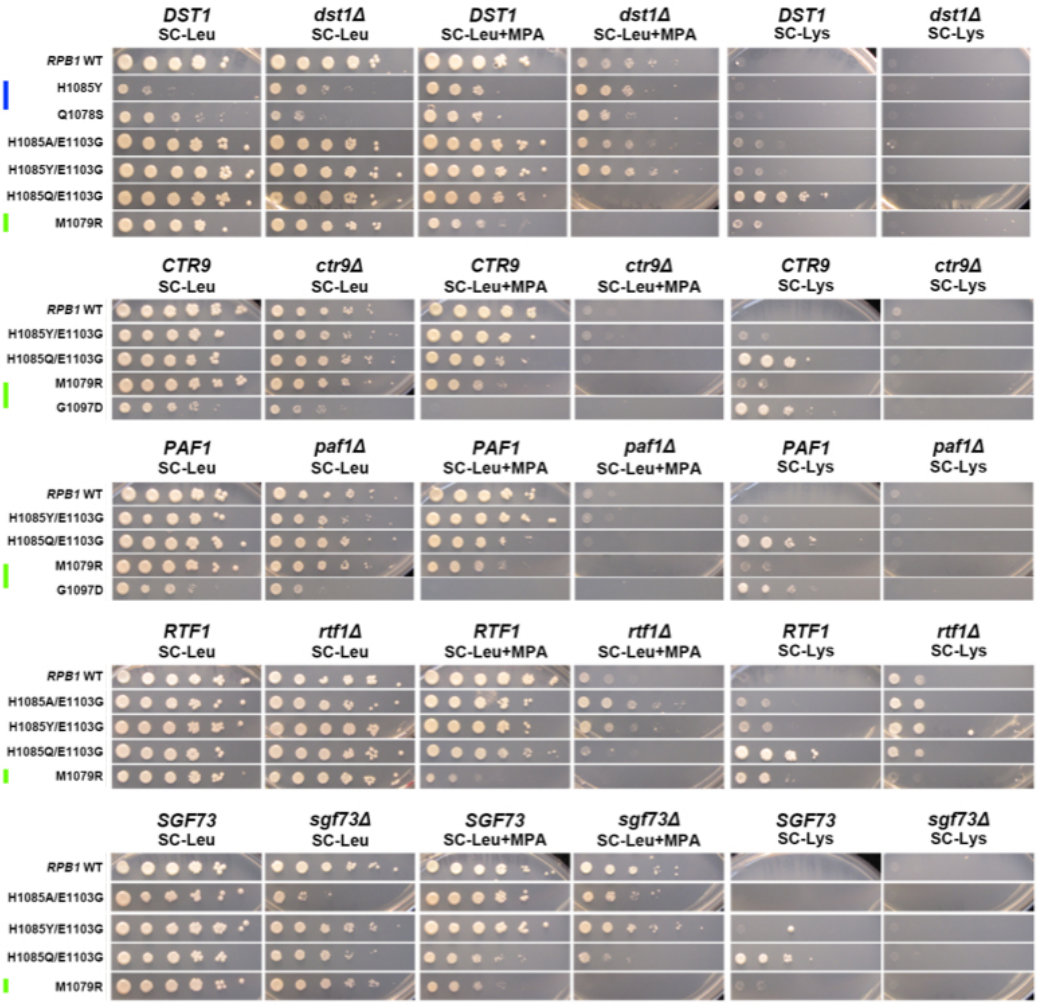


Figure 3-9 Continued.

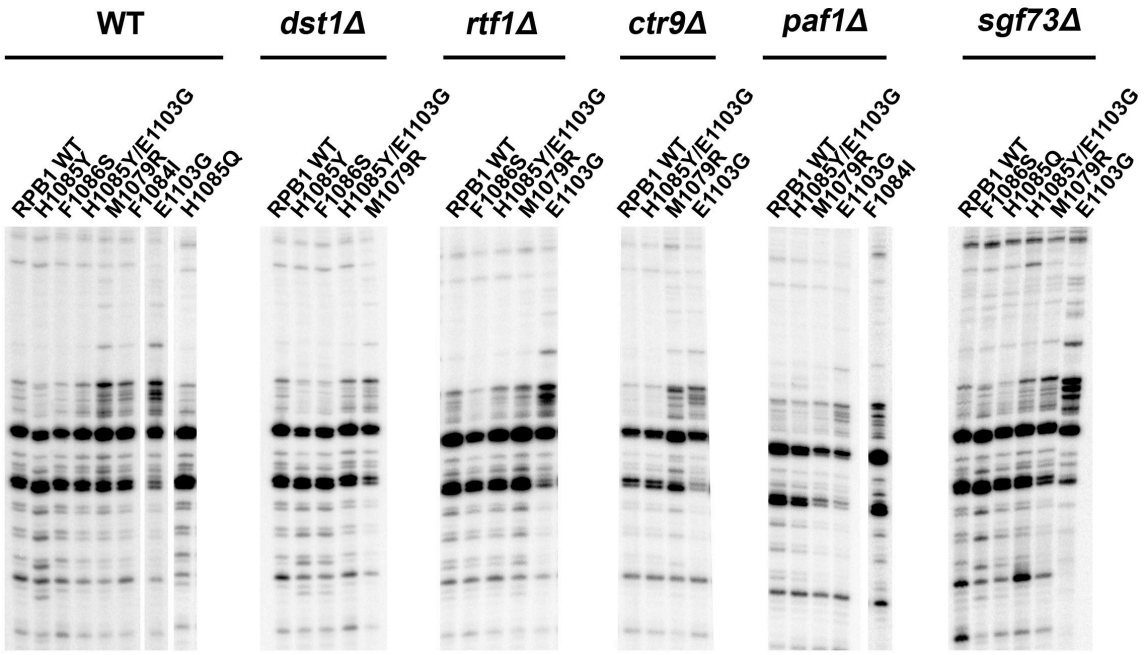


Figure 3-10. Transcription start sites of genetic interactor deletions at *ADH1* detected by primer extension.

One representative experiment of at least three independent replicates is shown. Genetic interactor deletions are labeled above each group of relevant lanes for the double mutant combinations of a particular gene deletion mutant with Pol II alleles. Labels for relevant *rpo21/rpb1* genotypes are above each lane. Quantifications of these experiments are shown in **Figure 3-8**.

Discussion

Combination of mutant alleles allows the relationships of different factors to be probed, with the hope of revealing distinctions between their contributions to different processes, and possibly suggesting mechanism. A number of factors contribute to TSS selection in *S. cerevisiae* and their relationships have been probed here to understand the requirements for normal Pol II initiation. Previous analyses showing that distinct classes of TSS mutant exist in yeast (upstream and downstream shifting) supporting a scanning model for identification of TSS. When mutants of differing TSS shift class have been combined (Freire-Picos et al., 2005; Ghazy et al., 2004; Kaplan et al., 2012; Sun and Hampsey, 1995), they generally exhibited mutual suppression of TSS defects coupled to suppression of growth defects, while mutants of the same class exhibited enhancement of TSS and growth defects.

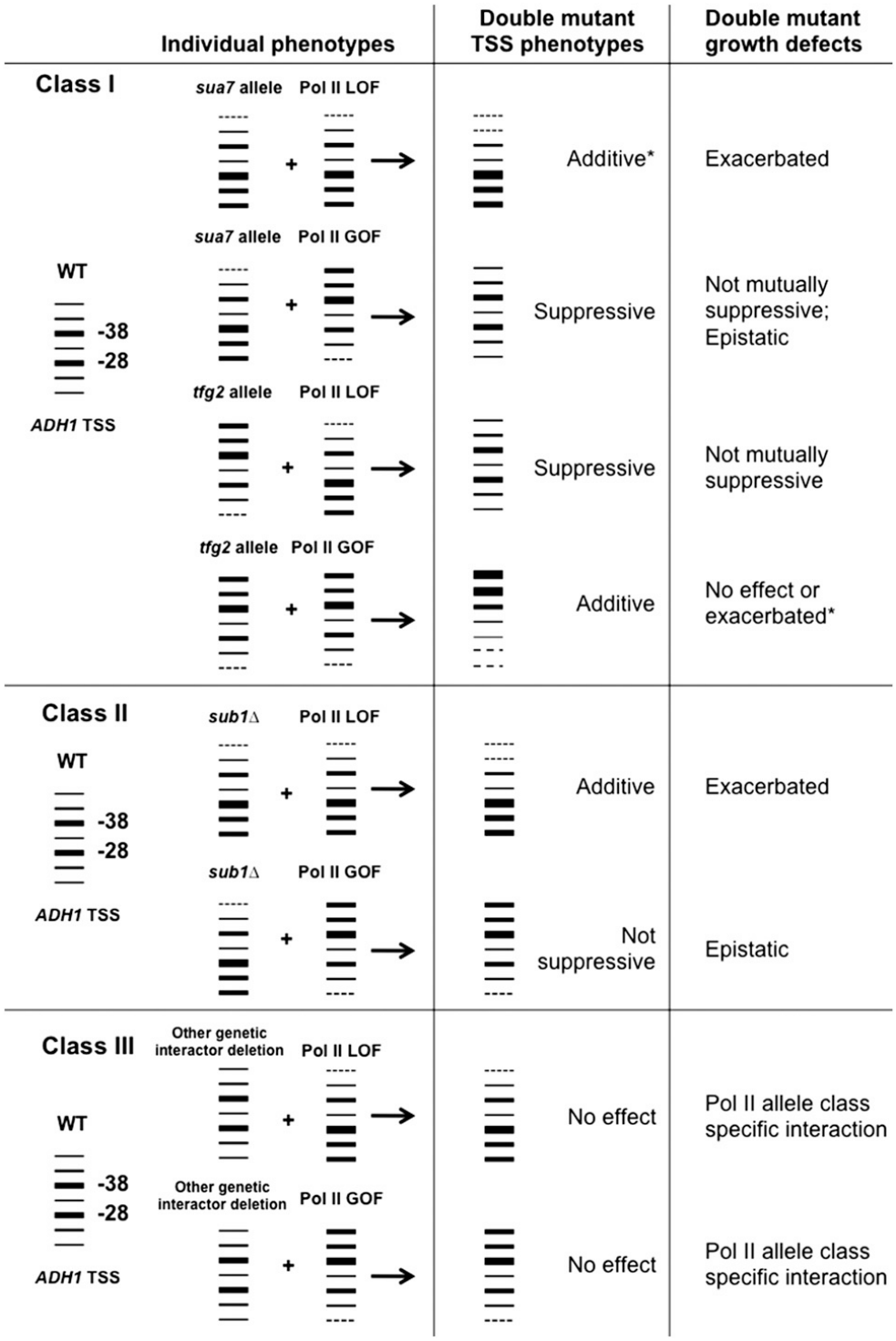
We previously discovered that Sub1 (homolog of PC4) has wide-ranging genetic interactions with Pol II mutants in a manner correlating with class of Pol II mutant (GOF or LOF)(Braberg et al., 2013). Sub1 was originally genetically isolated as a high copy suppressor of TFIIB alleles and biochemically as a positive transcription factor that stimulates basal transcription (Henry et al., 1996; Knaus et al., 1996). Deletion of *SUB1* causes synthetic lethality in combination with *sua7* TSS defective mutants (Knaus et al., 1996). The strong genetic interactions between *SUB1* and *SUA7* suggested a close association of their function in initiation and TSS selection and *in vivo* growth. We found that *sub1*Δ caused *ADH1* TSS distribution to shift downstream, and exacerbated the downstream shifts of Pol II LOF mutants, correlating with exacerbation of growth

defects in *sub1*Δ-Pol II LOF mutant double mutants (Braberg et al., 2013). Distinct from the general trend conferred by combination of TSS shifting alleles mentioned above, *sub1*Δ TSS effects were not additive with Pol II GOF mutants, instead epistasis was observed for Pol II GOF alleles combined with *sub1*Δ (Braberg et al., 2013). This epistasis was in contrast to *sub1*Δ enhancement of Pol II LOF alleles for both growth phenotypes and TSS shifts. In light of these different classes of relationships previously observed among mutants altering TSS selection, we examined the relationships between GTF alleles and Pol II alleles with altered trigger loops, and relationships between these same Pol II TL alleles and other Pol II genetic interactors.

We found that, unlike previous combinations of TSS shifting alleles, suppression of TSS defects by Pol II mutant-GTF combinations could be partially separated from their effects on growth. Furthermore, we found that previous observations regarding *sub1*Δ's relationship with Pol II alleles for TSS determination were relatively unique compared to a number of other Pol II genetic interactors examined here. Taken together, we can now discern at least three types of genetic relationships among TSS-altering alleles, Pol II interacting alleles, and Pol II active site mutants, in regards to double mutant modulation of TSS and growth defects (**Figure 3-11**). First, GTF and Pol II mutants that each alters start sites on their own, have additive or suppressive effects on TSS distribution at *ADH1*, depending on the nature of their single mutant defects (Class I). Unlike combinations of GTF alleles and a subset of previously examined *rpb* alleles, TSS defect suppression is partially uncoupled from growth defect suppression for oppositely acting TSS alleles. Negative growth interactions between GTF alleles and Pol

II alleles correlated with exacerbated TSS defects, but were of much greater strength for downstream shifting GTF TSS alleles and Pol II LOF alleles than for upstream shifting GTF TSS alleles and Pol II GOF alleles. These observations suggest that *S. cerevisiae* growth is much more sensitive to defects in initiation that result from decreased initiation efficiency. Second, *sub1*Δ is thus far unique in the strong epistasis observed between Pol II GOF alleles and *sub1*Δ for TSS selection at *ADH1* (Class II). This indicates that *sub1*Δ defects are distinct from those of TFIIB alleles, and that while *sub1*Δ appears to be bypassed in Pol II GOF alleles, *sua7* defects are not, though there is some observable suppression of specific phenotypes in *sua7*-Pol II GOF mutant strains. Third, other tested genetic interactors with Pol II do not generally have TSS defects on their own or modulate Pol II TL mutant TSS defects (Class III), suggesting that the strong correlation between TSS defects and a very broadly Pol II interacting mutant, *sub1*Δ, is relatively unique. Genetic interactions we observe between these factors and Pol II mutants may be originated (caused) by other transcriptional defects.

Figure 3-11. Model of relationships between TSS distribution shifting GTF alleles, Pol II active site alleles and other genetic interactors.
GTF alleles and Pol II mutants have additive or suppressive effects on TSSs at *ADH1* while showing exacerbation when same direction TSS mutants are combined but no mutual suppression of growth defects when opposite polarity TSS mutants are combined (Class I); *sub1*Δ has enhancement or epistasis with Pol II alleles on both TSS defects and growth defects, which is thus far unique among tested mutants (Class II); while other tested genetic interactors do not modulate TSS defects on their own or in combination with Pol II alleles yet exhibit a wide range of genetic interactions with Pol II TSS shifting mutants, suggesting relationships based on defects outside of TSS selection or initiation (Class III). *: Description based on viable double mutants.


Individual phenotypes
Double mutant TSS phenotypes
Double mutant growth defects
Class I
sua7 allele
Pol II LOF
Additive*
Exacerbated
sua7 allele
Pol II GOF
WT
-38
-28
ADH1 TSS
Suppressive
Not mutually suppressive; Epistatic
tfg2 allele
Pol II LOF
Suppressive
Not mutually suppressive
tfg2 allele
Pol II GOF
Additive
No effect or exacerbated*
Class II
sub1Δ
Pol II LOF
WT
-38
-28
ADH1 TSS
Additive
Exacerbated
sub1Δ
Pol II GOF
Not suppressive
Epistatic
Class III
Other genetic interactor deletion
Pol II LOF
WT
-38
-28
ADH1 TSS
No effect
Pol II allele class specific interaction
Other genetic interactor deletion
Pol II GOF
No effect
Pol II allele class specific interaction

While our experiments suggest that Pol II activity-dependent growth defects can be uncoupled from observed TSS defects, open questions remain concerning the mechanisms by which Pol II start sites are determined. RNA polymerases prefer to initiate at YR (−1, +1) sites, where the initiating nucleotide of an RNA is a purine, with a pyrimidine just upstream. A crystal structure of a viral RNA polymerase suggested that this sequence preference was likely due to purine stacking between the initiating NTP and a purine at the −1 position on the template strand (meaning a pyrimidine at −1 on the transcribed strand), and a set of very recent bacterial RNAP structures confirm this for multisubunit RNAPs (Basu et al., 2014; Gleghorn et al., 2011; Zhang et al., 2014). It seems likely that there are additional sequence determinants controlling efficiency of usage of any particular start site. For example, at *ADH1* transcription mutants appear to alter the probability of usage of YR sequences that are used by WT present in the start region. While Pol II GTFs might be positioned within the Pol II PIC to interact with sequences and "read" for the start site, the primary determinants for TSS preference are the −1/+1 bases that are located deep in the active site and not bases those juxtaposed to hypothetical or predicted GTF locations in the PIC (these would mostly be upstream sequences). Moreover, the types of TSS changes observed for GTF mutants are phenocopied by mutations in the Pol II trigger loop, suggesting they might arise from similar types of defects in transcription. What could these similar defects be?

The altered patterns of starts observed in most upstream or downstream shifting start site mutants appear to be stereotypical to each class, meaning the positions of starts that are more likely used at *ADH1* in mutants are the same within each mutant class as

defined by upstream or downstream shifting, not by whether they are in GTF subunits or the Pol II active site. However, the fraction of usage on these usable start sites differs (strong or weak shift in distribution of TSS usage) depending on how much an allele deviates from WT catalytic activity for Pol II alleles, or how strong they appear genetically (for GTF alleles). In other words, different mutants may not necessarily alter initiation sequence *preference*, but instead shift the initiation *probability* of use in a polar fashion within some set of already usable start sites. In this view, Pol II initiation efficiency may cooperate with a directional scanning process that has its own rate. Pol II with increased catalytic activity enables initiation earlier within the scanning window, increasing catalytic efficiency (initiation probability) of earlier usable start sites usages, shifting TSS usage distribution upstream.

Additive effects on *ADH1* TSS distributions in Pol II-GTF mutant strains indicate that individual defects of each allele are present in the double mutant strains. How might we understand these defects? Defects in *sua7* are consistent with defects in initiation efficiency and likely represent reduced TFIIB functions. TFIIB function in concert with the Pol II active site might represent communication between TFIIB and the active site as has been proposed (Sainsbury et al., 2013), or parallel roles for TFIIB and the Pol II active center during putative TSS scanning. We speculate that *sua7* Pol II LOF double mutants that have severe growth defects or are lethal as shown in Figure 2 have exacerbated defects in initiation efficiency from those of single mutants, *e.g.* those shown for *sua7-1* (Cho and Buratowski, 1999). Furthermore, mutual suppression of TSS defects in GTF-Pol II double mutants is predicted to result from suppression of initiation

defects, which might be tested for GTF-Pol II mutant combinations in GTF-dependent biochemical systems for abortive or productive Pol II initiation such as the system employed recently by Fishburn and Hahn (Fishburn and Hahn, 2012). Because Pol II alleles are expected to have additional defects in elongation and termination, GTF alleles do not strongly suppress overall growth defects of Pol II mutants even though TSS defects in some cases examined here are strongly suppressed. Our previous work detecting splicing defects as a consequence of Pol II alleles' presumed altered elongation functions did not detect similar splicing defects for GTF alleles. Defects detected were milder and of opposite polarity for *sua7-3* and *tfg2*Δ*261-273* strains relative to Pol II alleles with upstream or downstream shifts in TSS distributions, arguing against predictable defects of these alleles in Pol II elongation (Braberg et al., 2013).

tfg2 TSS phenotypes are similar to those of Pol II GOF alleles and raise the question of how alteration of TFIIF function alters TSS distribution. Previous work had indicated that a *TFG1* mutant exhibited an increased ability to stimulate Pol II activity in an abortive initiation assay (Khaperskyy et al., 2008). Fishburn and Hahn recently reported a negative role for TFIIF in suppressing TSS usage at upstream positions of promoters (nearer to a TATA element) (Fishburn and Hahn, 2012). Such a negative role is likely balanced by positive requirements for TFIIF activity in promoting initiation. Conceivably, *tfg1* and *tfg2* mutants have this negative role– possibly an autoinhibitory function of TFIIF that is alleviated during scanning to downstream positions– specifically or selectively compromised. In light of such a model, upstream TSS shifting *tfg* alleles would confer increased initiation activity, phenocopying Pol II GOF mutants

for altered TSS distribution. In *in vitro* transcription experiments, TFIIF stimulation of abortive initiation was compromised by the Pol II LOF H1085Y allele of the TL (Cabart et al., 2014). TFIIF stimulation of abortive initiation likely represents one of a number of positive TFIIF roles in initiation. If the observed *in vitro* stimulation were related to upstream TSS shifts at *ADH1* observed in *tfg* alleles, compromise of the TL might be expected to be epistatic to *tfg* phenotypes based on biochemical results. As *tfg2* alleles can still alter TSS distribution in Pol II LOF alleles such as H1085Y, it suggests that the biochemical requirement *in vitro* of TFIIF for a WT TL may be distinct or bypassed by *tfg2* alleles studied here.

The mechanism driving *S. cerevisiae* start site scanning likely derives from a combination of factors, Pol II activity and transcription bubble opening promoted by TFIIH. Assuming that upstream bubble opening, as was observed at *GAL1* and *GAL10* (Giardina and Lis, 1993), is universal at yeast promoters, a major question is how does the bubble transit to the distal start sites? One possibility is that a large transcription bubble is extended to the start site region. In this case, PICs would need to accommodate extensive single-stranded DNA. A recent cryo-EM structure and model of the yeast PIC appears consistent with such accommodation (Murakami et al., 2013). In this instance, TFIIH might drive extension of the downstream bubble edge through the activity of its Ssl2/Rad25 subunit (the yeast homolog of human ERCC3/XPB). An alternative model, where a smaller region of melted DNA translocates along with the open PIC towards the start region may also be possible, but almost nothing is known of the organization of nucleic acids in such hypothetical complexes or how translocation of the putative bubble

would be controlled during initiation. In translocating Pol II elongation complexes, GTFs are not present and interactions between Pol II, both DNA strands, and nascent RNA organize the upstream edge of the transcription bubble.

Genetic analyses have allowed us to examine the relationships between Pol II activity mutants, known initiation factors, and candidates for possible modifiers of Pol II initiation activity. Our experiments indicate that Pol II genetic interactors need not perturb TSS selection, and that initiation defects are likely only a partial driver of Pol II allele growth phenotypes. Altered Pol II activity through TL defects do not bypass or appear epistatic to the alleles of TFIIB or TFIIF studied here for *ADH1* TSS selection, suggesting that they each function separately as part of a concerted process to promote efficient TSS selection.

CHAPTER IV

CHANGES IN POL II CATALYTIC ACTIVITY INFLUENCE TRANSCRIPTION START SITE UTILIZATION AND OTHER TRANSCRIPTION CONTROLS GENOME WIDE

Disclaimer

Chapter IV is work not yet published, I will be first author when it is submitted for publication.

TSS sequencing: We conducted genome wide 5′ capped RNA mapping to detect transcription start site of mRNA using SOLiD sequencing and mapped GTFs using ChIP-Exo sequencing. I performed RNA preparation and purification, and the Nickels lab (Rutgers University) constructed sequencing libraries and conducted SOLiD 5′-RNA sequencing. Dr. Scott Schwartz (TAMU Genomic center, now at the University of Texas) conducted bioinformatics analyses using in-house scripts with raw sequencing reads. Dr. Schwartz performed most of upstream bioinformatics analysis for TSS with his in-house pipeline in close consultation with me and Dr. Kaplan. I conducted a section of bioinformatics related to initiator motif usage analysis in yeast and other organisms using a pipeline and script I built inspired by Dr. Schwartz's scripts. I performed all downstream analyses including heatmaps, TSS comparison by promoter classes, aggregation analyses, sequence motif analyses, gene expression correlation analyses within our data and comparison with other published data, promoter architecture based

distance and expression analyses. I received experimental assistance from lab member Indranil Malik and previous lab member Ping Cui in RNA preparation for a few promoter constructs. All northern blot experiments were performed by Indranil Malik.

ChIP-exo sequencing: For the first round of ChIP-exo experiment, a previous lab member Dr. Pavel Čabart constructed epitope-tagged GTFs in Pol II mutant strains, prepared chromatin samples and sent them to the Pugh lab (PSU); the Pugh lab performed ChIP-exo treatment, library construction, and sequencing. Currently, we are attempting to obtain a second repeat of ChIP-exo samples in collaboration with Pugh lab. ChIP-exo data used in this dissertation are from first round of ChIP-exo. Dr. Schwartz conducted the first rounds of bioinformatics analyses with in-house scripts and pipelines designed to specifications of Dr. Kaplan and myself. I adapted and modified Dr. Schwarz's scripts and conducted bioinformatics analyses and downstream data analyses for the ChIP-exo data included in this chapter.

Nucleosome mapping: I performed MNase experiments on WT and two other Pol II alleles and purified mononucleosomal DNA. The TAMU/Agrilife genomics center performed paired end Illumina sequencing on this DNA; raw data were processed by Jordi Abante Llenas (TAMU Department of Electrical and Computer Engineering). I have performed all downstream data analyses.

Summary

Transcription initiation by RNA Polymerase II (Pol II) is an essential step in gene expression and can be highly regulated. Promoter recognition and transcription start site (TSS) selection by Pol II require the activity of a number of conserved factors. *Saccharomyces cerevisiae* (*S. cerevisiae*) Pol II has been proposed to scan the promoter region directionally for TSSs during the process of TSS selection. If and how different promoter classes differ in this process are unknown. Mutations in initiation factors and the Pol II active site have been shown to alter TSS selection/utilization at model genes. To investigate effects of altered Pol II activity on TSS distribution we performed genome wide detection of presumptive TSSs by SOLiD sequencing of capped RNA 5′ ends. We show that Pol II activity mutants and General Transcription Factor (GTF) TFIIB and TFIIF mutants confer TSS shifts globally at all classes of promoter in yeast and we explore two models that may explain the shifts- a "motif preference" model and an "altered efficiency window" model. We have investigated how distinct promoter classes shape initiation differently by exploring their TSS characteristics (distances to recognizable promoter elements, broadness of TSS distributions) and the sequence contexts within promoter regions. We observed changes to promoter initiation properties dependent on promoter architecture including magnitude of TSS shifts and changes to overall expression levels in Pol II activity mutants. We also examined functional requirements for alternative promoter elements at core promoters lacking consensus TATA elements. Finally, we investigated role of TSS selection in nucleosome positioning. We presented evidences that +1 nucleosome positioning correlates with

GTF localization, which is a proxy for presumed PIC assembly. We showed that positioned nucleosomes are shifted directionally in the same direction as the shift of TSSs in a slow Pol II mutant, in a manner sensitive to promoter expression level. Furthermore, we utilized a bioinformatics approach to examine metazoan TSS usage characteristics and compare them to TSS distributions in yeast to determine level of similarity and compatibility with a scanning mechanism.

Methods

Yeast strains

Yeast strains used in this study were constructed as described previously (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012). Briefly, plasmids containing *rpo21/rpb1* mutants were introduced by transformation into a yeast strain containing a chromosomal deletion of *rpo21/rpb1* but with wild type *RPO21/RPB1* on a plasmid with a selectable marker. GTF mutant parental strains used for GTF single or GTF/Pol II double mutant analyses were constructed by chromosomal integration of GTF mutants into their respective native locus by way of two-step integrations. Strains used in ChIP-exo were TAP-tagged at target genes using homologous recombination of the TAP tag (Puig et al., 2001) and a selectable marker at the 3′ ends of appropriate genes. Strains with *rpb1* mutants were generated by plasmid shuffling (Boeke et al., 1987) to select for cells containing the *rpb1* mutant plasmids in absence of the *RPB1* WT plasmid. Yeast strains in all experiments were grown in YPD rich media unless noted otherwise.

SoLID 5′-RNA sequencing

Yeast strains were grown to reach mid-log phase (~1.5×10^7/ml) in rich media (YPD) and harvested, having been first diluted from a saturated overnight YPD culture. Total RNAs were extracted by a hot phenol-chloroform method (Schmitt et al., 1990), followed by on-column incubation with DNase I to remove DNA (RNeasy Mini kit, Qiagen), and processing with a Ribo-Zero rRNA removal kit (Epicentre) to deplete rRNA. To construct the cDNA library, samples were treated with terminator 5′ phosphate-dependent exonuclease to remove RNAs with 5′ monophosphate (5′ P), and remaining RNAs were purified. Tobacco acid phosphatase (TAP, Epicentre) was added to convert 5′ PPP or capped RNAs to 5′ P RNAs, and RNAs were purified. A SOLiD 5′ adaptor was ligated to RNAs with 5′ P (this step excludes 5′ OH RNAs), followed by gel size selection of 5′ adaptor ligated RNAs and reverse transcription (SuperScript III RT, Invitrogen) with 3′ random priming. RNase H (Ambion) was added to remove the RNA strand of DNA-RNA duplexes, cDNA was size selected for 125-500 nt lengths. These cDNA libraries were amplified using SOLiD total RNA-seq kit and SOLiD Barcoding kit, final DNA was gel size selected for 160-300 nt length, and sequenced by SOLiD (Applied Biosystems) as described previously (Goldman et al., 2011).

ChIP-exo sequencing

ChIP-exo experiments were performed as described previously (Rhee and Pugh, 2011). Briefly, yeast strains were amplified to mid-log phase (~1.5×10^7/ml) in rich medium (YPD) from a saturated overnight and were crosslinked with formaldehyde (1% final concentration) for 20 min and then quenched by 0.25M glycine. Cells were washed,

lysed with glass beads by beat beating (30s on 60s off, 7-8 rounds; checked under microscope) followed by shearing of chromatin to ~200-500 nt by sonication at 4 ˚C (Diagenode Bioruptor, 30 s on 30 s off, high power, 18 cycles or until chromatin was in the right size range) in FA-lysis buffer without detergents. Solubilized chromatin was Immunoprecipitated using IgG-bound sepharose resin and washed. Immunoprecipitated chromatin on resin was end polished by T4 DNA polymerase (NEB). Adaptor sequences were then ligated on both ends of linker DNA, followed by nick repair using phi29 polymerase (NEB). λ exonuclease (NEB) was used for 5′ → 3′ digestion of sonicated protein-dsDNA and RecJ exonuclease for 5′ → 3′ digestion of ssDNA to minimize background. Resin was washed, DNA was eluted with TEV protease (Invitrogen), and reverse crosslinked by incubating at 65 ˚C with Proteinase K (Roche). DNA was extracted by phenol:chloroform:isoamyl alcohol, denatured to ssDNA (each overhang of dsDNA was different and corresponded to one border of protein binding), amplified to dsDNA with oligos with -OH on both ends to amplify exonuclease treated strand only. Adaptors are ligated only to λ exonuclease digested ends of dsDNA, ligated DNA was then amplified, size selected for 120−160 nt and are sequenced by SOLiD sequencing (Applied Biosystems)

Nucleosome MNase sequencing

Nucleosomal DNAs were prepared by a method described elsewhere (van Bakel et al., 2013) but with the following modifications. Yeast strains were grown in rich medium (YPD) to mid-log phase (~1.5×10^7/ml) and cross-linked with methanol-free formaldehyde (1% final concentration, Polysciences Inc) for 30 min and quenched with

0.25M final concentration of glycine (from 2.5M stock, pH=7). Cells were washed and digested with zymolyase-20T (Sunrise International) (6mg for 500ml culture) for ~17 min or until ~90% cells appeared as spheroplasts, followed by MNase (Thermo Scientific) digestion with different levels of MNase to generate "less" and "more" digested nucleosomes (in general digests were limited such that at least mono, di, and trinucleosomes were still apparent in the sample). Crosslinks on nucleosomes were reversed at 65 ˚C in the presence of Proteinase K (G-Biosciences) overnight, DNA was extracted by phenol/chloroform, and digested with RNase A (Thermo Scientific) to remove RNAs. Nucleosomal DNA was separated on 1.5% agarose gels containing SYBR gold dye (Life Technologies) and mono-nucleosome bands were identified and selected under blue light and gel purified (Omega Biotek). Mono-nucleosomal DNA was sequenced with paired-end Illumina Tru-Seq sequencing.

Bioinformatics pipelines

TSS and nucleosome sequencing reads were aligned uniquely to SacCer3 (v64) reference genome using the Bowtie sequencing aligner (Langmead et al., 2009) allowing 2 (TSS sequencing) and 1 (nucleosome sequencing) mismatches; ChIP-exo reads were aligned to v56 reference genome by Corona Lite software provided by SOLiD allowing up to 3 mismatches. Only uniquely aligned reads were kept. ChIP-exo reads alignments were converted to SacCer3 (v64) genome alignment using the LiftOver tool (http://genome.ucsc.edu/). Strand-conscious starting nucleotides of TSS and ChIP-exo and midpoints of nucleosomes were determined for each aligned read. Promoter windows were selected by expanding 200 nt up and downstream from centered

TATA/TATA-like elements proposed by (Rhee and Pugh, 2012) (408 nt total). TSS-centered promoter windows were generated by determination of TSS of 50th percentile in each promoter in WT reads derived from 446, 456, 497, 499 libraries; see below) and expanding 250 nt upstream and 150 nt downstream from the position (401 nt total).

Results

Detection of TSS genome wide by SOLiD 5′-RNA sequencing

We performed SOLiD (Sequence by Oligonucleotide Ligation and Detection) deep sequencing on capped mRNAs to determine 5′-RNA ends and map presumptive transcription start sites (TSSs) on a global scale in *S. cerevisiae*. SOLiD is an adaptor ligation based high throughput sequencing method that can be used to identify the 5′ ends of transcripts (Goldman et al., 2011). SOLiD has become a popular choice for deep sequencing due to high throughput/high accuracy among existing next generation sequencing methods; its disadvantage is that it is relatively slow, and is better for short read sequencing, and is now being outstripped in throughput by Illumina platforms (Liu et al., 2012; van Dijk et al., 2014). **Figure 4-1** describes the overall flow of TSS mapping, determination of TSS properties, and TSS shifts within specified promoter windows. Also see **Methods**.

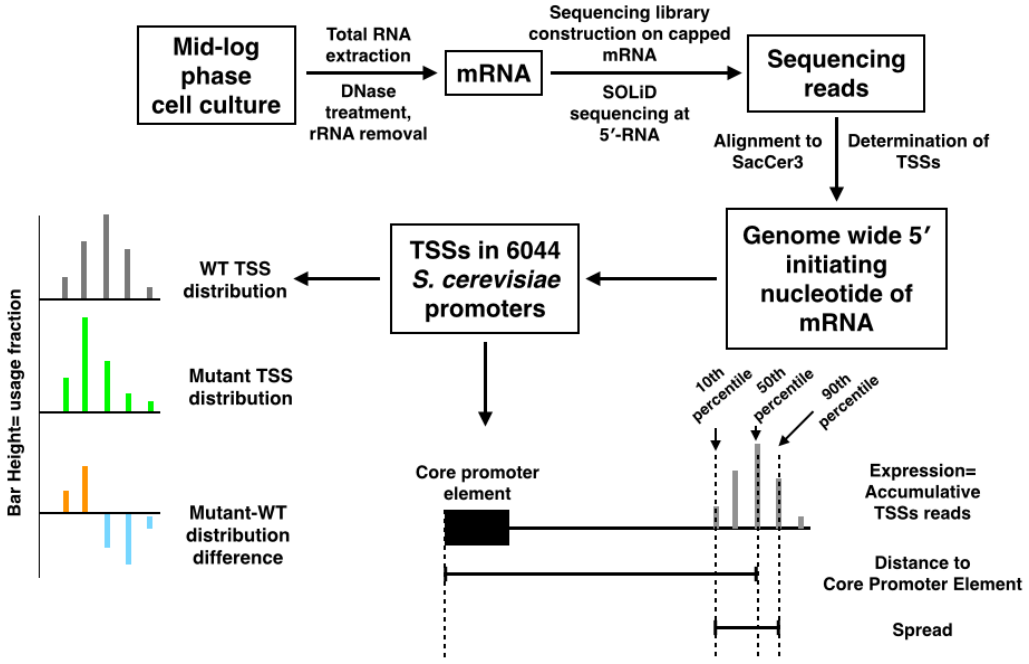


Figure 4-1. Experimental pipeline of SOLiD 5′ RNA sequencing.
Yeast cultures were harvested at mid-log phase, total RNAs were extracted, followed by DNase digestion and rRNA removal. Sequencing libraries enriching capped mRNAs were constructed and were sequenced by SOLiD (Applied Biosystems). Reads were mapped to SacCer3 reference genome, followed by determination of TSSs for each promoter. TSSs found in designated promoter windows are examined for following characteristics. Expression: Accumulative TSS reads in promoter window; TSS distance to core promoter element: Distance between median TSS position and predicted core promoter element or proxy estimate of core promoter; TSS spread: Distance between 90^{th} percentile TSS and 10^{th} percentile TSS in each promoter. Changes of TSS distribution conferred by a mutant are determined by subtracting WT normalized TSS density in window of interest from mutant normalized TSS density.

Previously we discovered that Pol II mutants that alter catalytic activity *in vitro* changed distribution of TSS utilization *in vivo* at a few model genes (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012). Shifts observed were similar to those observed for different GTF mutant classes ((Eichner et al., 2010; Pinto et al., 1994; Sun et al., 1996), see note about *sua7-58A5* in Chapter III). Fast/superactive Pol II mutants that confer faster elongation *in vitro*, *tfg2* (a subunit of TFIIF) Δ*146-180* and *tfg2*Δ*146-180*/E1103G showed upstream shift in TSS distribution while slow Pol II mutants and *sua7-58A5* (an allele of *SUA7*, the gene encoding TFIIB) showed opposite changes in TSS distributions – a downstream shift in TSS distribution at model genes. Our observations of TSS distribution alterations by GTF mutants are consistent with previous studies. At the model gene *ADH1*, the magnitude of shifts in both slow and fast mutant classes correlated with their severity of activity changes, growth phenotype indicative of transcriptional defects as well as genetic interactions (Chapter II, III(Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012)). Based on these observations, we wished to determine how these Pol II activity/GTF mutants altered TSS selection/utilization and other initiation properties in genome wide.

We mapped TSSs for 38 samples in total, in two batches of sequencing, comprising four repeats of *RPO21*/*RPB1* WT, nine *rpo21/rpb1* mutants that alter Pol II catalytic activity (fast or slow) in at least duplicates, two GTF single mutants in duplicates – *sua7-58A5* (TFIIB) and *tfg2*Δ*146-180* (TFIIF) – and one *tfg2*Δ*146-180* (TFIIF)/E1103G (fast Pol II) double mutant in duplicate. To analyze effects on TSS selection in our mutants, we focused on regions of the genome expected to harbor

promoters. We chose 6044 promoters (**Table 4-1**) that were classified previously into promoter classes by their initiation factor binding (Taf1-enriched or depleted) and presence of promoter elements (TATA or TATA-like) (Rhee and Pugh, 2012). We designated TATA/TATA-like element centered promoter windows by assaying sequence 200 nt upstream and downstream relative to the proposed TATA/TATA-like elements ("promoter window", 408 nt total range). On average, we obtained ~10 million uniquely mapped reads per library, of which ~75% of reads were within our 6044 promoter windows (percentage of strand specific promoter reads in reads genome wide on both (forward and reverse) strands). Promoter enrichment for sequencing reads was greater than 70 fold when measuring reads per nucleotide compared to reads genome wide fitted to a Poisson distribution, which assumes an even distribution of reads at all nucleotide positions (Haight, 1967)(**Table 4-2**). This enrichment indicates that our TSS protocol is highly efficient for identifying presumptive capped RNA 5′ ends and is not dominated by RNA degradation products, which would be expected to have a wider distribution across the genome. Additional support that our data reveal *bona fide* TSSs is that we observe very high enrichment for $Y_{-1}\underline{R}_{+1}$ as the sequence motif for the 5′ end positions of our RNAs (Y=pyrimidine at the −1 position, *i.e.*, the immediate upstream position of the terminal nucleotide at the presumptive TSS; R=purine at the TSS position designated as $\underline{+1}$). $Y_{-1}\underline{R}_{+1}$ was previously identified as the initiator motif for *S. cerevisiae* and other eukaryotic Pol II enzymes (Hahn et al., 1985; Hampsey, 1998b; Kuehner and Brow, 2006; Zhang and Dietrich, 2005). In a representative WT library, the enrichment for $Y_{-1}\underline{R}_{+1}$ reads in promoter window is ~20 fold over reads mapping to the genome outside

of our promoter regions ($Y_{-1}\underline{R}_{+1}$ reads on both forward and reverse strand in promoter window / YR reads on both strands in rest of the genome, **Figure 4-2**). The evidence above supports that the 5′-RNA reads in our promoter windows are mostly detecting capped RNAs with expected characteristics of TSSs. TSSs for our model gene *ADH1* detected in our sequencing are consistent with TSSs *in vivo* detected by primer extension for both WT and TSS shifting mutants (Chapter II, III) (**Figure 4-3**).

Table 4-1 Promoters in this study

Promoter classification	Promoter class	Numbers of promoters	Misannotated TATA elements in the promoter class
TATA-dependent	TATA-containing	1011	120
	TATA-less	5033	132
Taf1-dependent	Taf1-enriched	4754	
	Taf1-depleted	1135	

Table 4-2 SOLiD 5′-RNA reads enrichment in promoter window region

Library #	WT/mutant Description	Promoter total reads	Unique reads	Promoter/ unique reads*	Promoter reads/nt enrichment**
446	RPB1 WT	9,919,395	13,159,120	75.38%	73.28
456	RPB1 WT	12,980,214	17,831,981	72.79%	70.77
447	L1101S	6,210,045	11,676,546	53.18%	51.70
457	L1101S	7,968,239	10,532,376	75.65%	73.55
448	G1097D	6,429,813	9,070,388	70.89%	68.92
458	G1097D	15,647,544	23,889,820	65.50%	63.68
459	H1085Y	18,708,852	26,915,227	69.51%	67.58
450	F1086S	11,050,097	14,477,490	76.33%	74.20
460	F1086S	14,127,286	18,953,069	74.54%	72.47
451	M1079R	14,202,763	17,786,521	79.85%	77.63
461	M1079R	7,723,806	9,648,992	80.05%	77.82
452	N1082S	5,677,990	9,328,417	60.87%	59.17
462	N1082S	14,547,777	20,363,467	71.44%	69.45
453	H1085Q	4,936,871	7,616,197	64.82%	63.02
463	H1085Q	15,835,789	21,322,912	74.27%	72.20
454	F1084I	8,612,066	13,494,293	63.82%	62.05
464	F1084I	14,853,609	19,853,378	74.82%	72.74
455	E1103G	6,693,084	9,065,769	73.83%	71.77
465	E1103G	11,193,890	13,928,954	80.36%	78.13
497	RPB1 WT	4,774,669	6,378,175	74.86%	72.78
499	RPB1 WT	7,257,937	9,220,893	78.71%	76.52
501	H1085Y	2,823,317	4,274,894	66.04%	64.21
503	H1085Y	2,409,371	3,864,558	62.35%	60.61
505	E1103G	5,963,317	7,312,890	81.55%	79.28
507	E1103G	5,463,759	6,802,625	80.32%	78.08
509	F1086S	5,652,627	7,230,051	78.18%	76.01
510	F1086S	4,276,209	5,459,023	78.33%	76.15
511	M1079R	9,293,054	11,679,490	79.57%	77.35
512	M1079R	4,283,993	5,102,430	83.96%	81.62
513	G1097D	1,786,312	2,287,034	78.11%	75.93
514	G1097D	3,908,110	4,869,871	80.25%	78.02
515	*sua7-58A5*	3,052,040	4,043,684	75.48%	73.38
516	*sua7-58A5*	1,867,421	2,354,301	79.32%	77.11
517	*tfg2Δ146-180*	5,038,321	5,927,347	85.00%	82.64
518	*tfg2Δ146-180*	6,750,150	8,027,127	84.09%	81.75

Table 4-2 Continued

Library #	WT/mutant	Promoter total reads	Unique reads	Promoter/ unique reads	Promoter reads/nt enrichment
519	*tfg2Δ146-180*/E1103G	4,319,222	5,507,545	78.42%	76.24
520	*tfg2Δ146-180*/E1103G	4,686,717	5,719,155	81.95%	79.67

Promoter/unique reads* : Promoter reads/ overall uniquely mapped reads
Promoter reads/nt enrichment**: (Promoter reads/ promoter window size (408 nt ×6044=2,465,952 nt)) ÷ (Unique reads/ genome size (24,314,21 nt))

Table 4-3 Reads in merged WT/mutant libraries in promoter windows

Merged Libraries	Mutation	Pol II activity	Reads
merged_446_456_497_499	WT		46,590,169
merged_447_457	L1101S	Fast	22,208,922
merged_448_458_513_514	G1097D	Fast	40,117,113
merged_450_460_509_510	F1086S	Slow	46,119,633
merged_451_461_511_512	M1079R	Fast	44,217,433
merged_452_462	N1082S	Slow	29,691,884
merged_453_463	H1085Q	Slow	28,939,109
merged_454_464	F1084I	Fast	33,347,671
merged_455_465_505_507	E1103G	Fast	37,110,238
merged_459_501_503	H1085Y	Slow	35,054,679
merged_515_516	*sua7-58A5*		6,397,985
merged_517_518	*tfg2Δ146-180*		13,954,474
merged_519_520	*tfg2Δ146-180*/E1103G		11,226,700

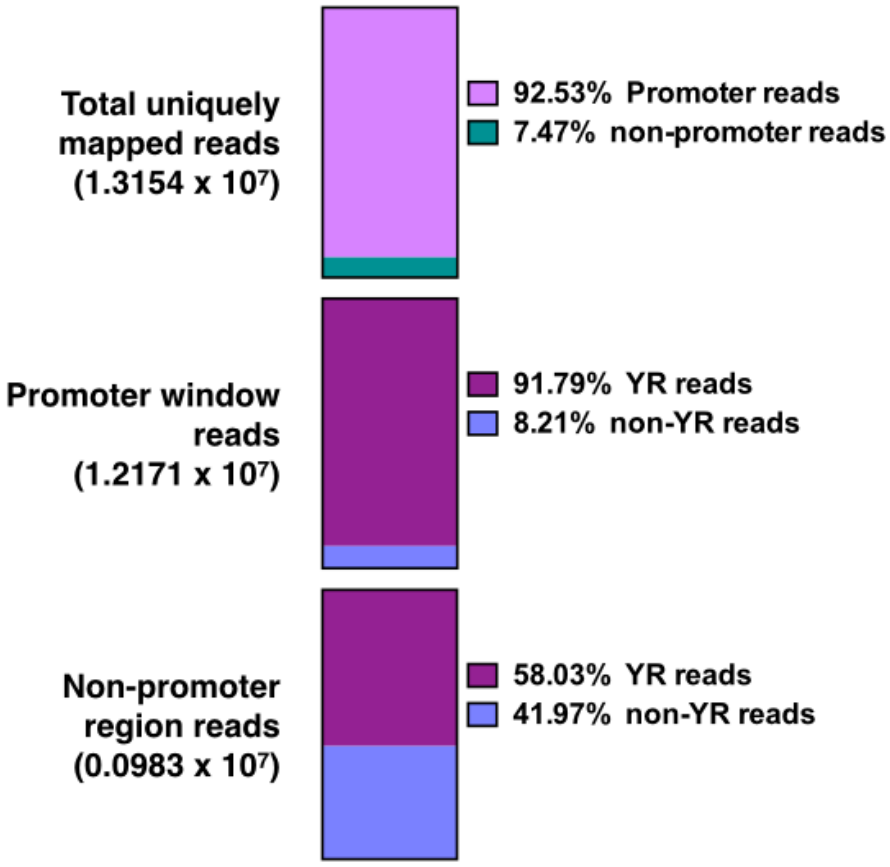


Figure 4-2. Enrichment of $Y_{-1}R_{+1}$ reads in promoter windows relative to the rest of the genome.
Reads (forward and reverse strands) in promoter (408 nt TATA/TATA-like centered) windows compared to total uniquely mapped reads genome wide in a representative WT (library 446) are shown. Reads that initiate with an R (A/G)) preceded by a Y (C/T) at the −1 position in the genome are counted as $Y_{-1}R_{+1}$ reads, all other reads are counted as non-YR reads. Note that our analyses likely lack some bona fide and therefore are probably an underestimate of read enrichment to actual promoters.

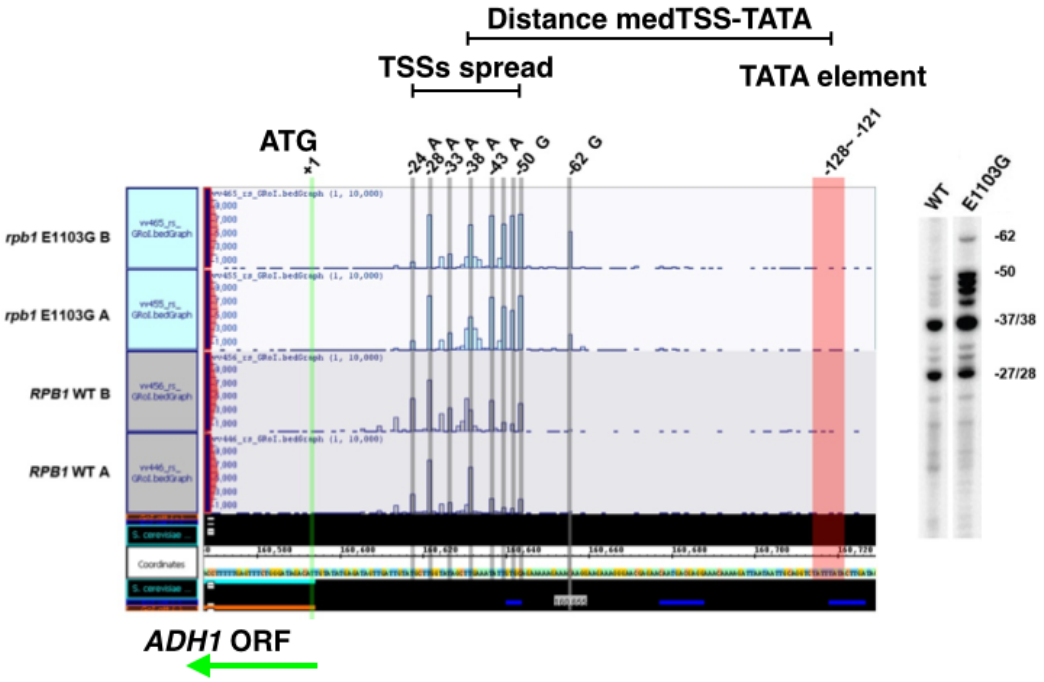


Figure 4-3. *ADH1* TSSs detected by SOLiD 5′ RNA sequencing are consistent with primter extension assay for WT and TSS shifting mutant.
Genome browser (IGB) view of SOLiD 5′ RNA sequencing data for the *ADH1* promoter region of WT and a upstream shifting mutant *rpo21*/*rpb1* E1103G is shown on the left. Two sequencing libraries (A and B) per WT/mutant are shown, with major start sites labeled as their relative position to the *ADH1* start codon ATG and starting nucleotide labeled on top. TSS parameters as calculated in **Figure 4-1** are shown. Primer extension assays for TSSs detection *in vivo* are for shown for WT and E1103G on the right.

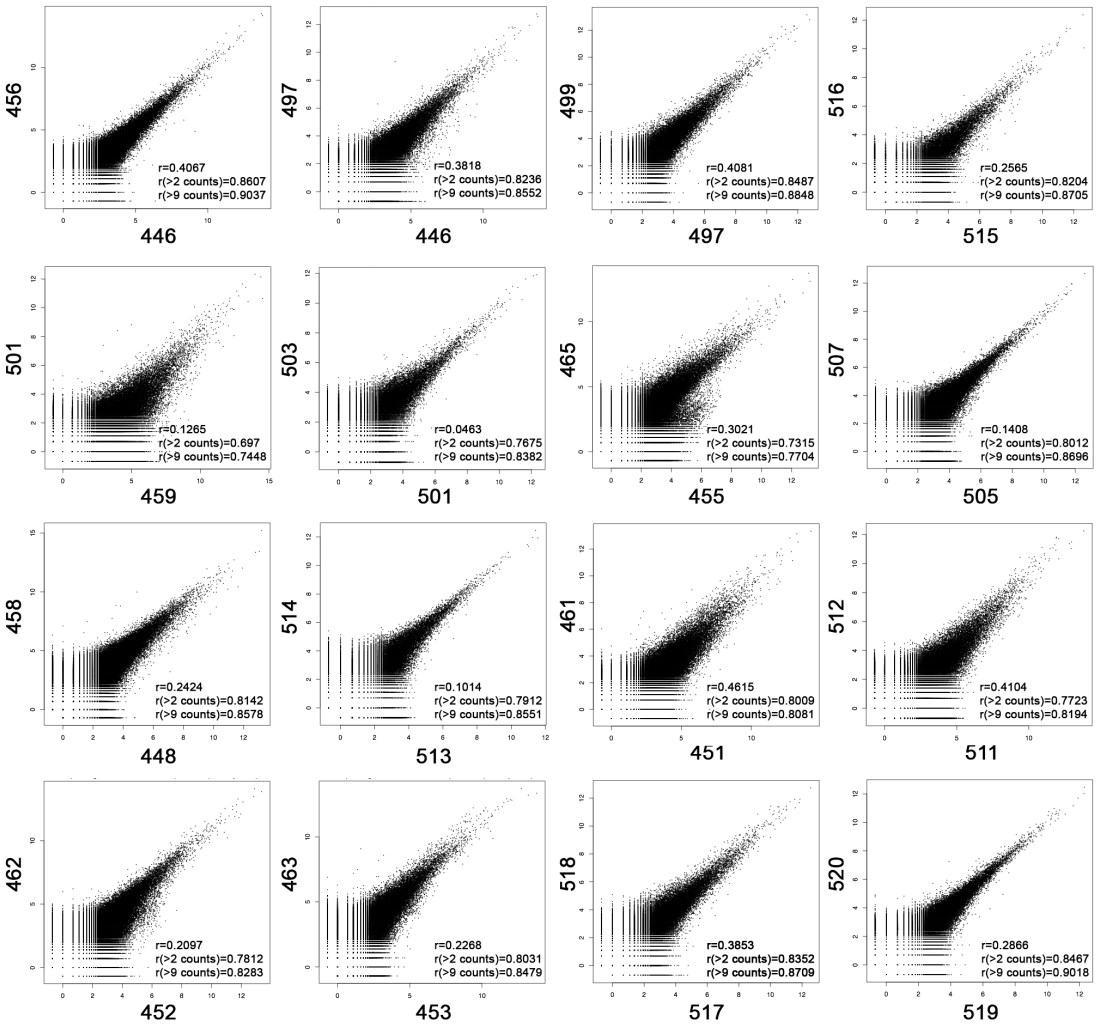


Figure 4-4. Position based correlations between replicate libraries are shown. See text for calculation. Library numbers (see **Table 4-2**) in comparison are labeled on axis. Pearson r for all data points, for data points above 2 read counts per position in all libraries and 9 read counts per position in all libraries are shown on top of each graph.

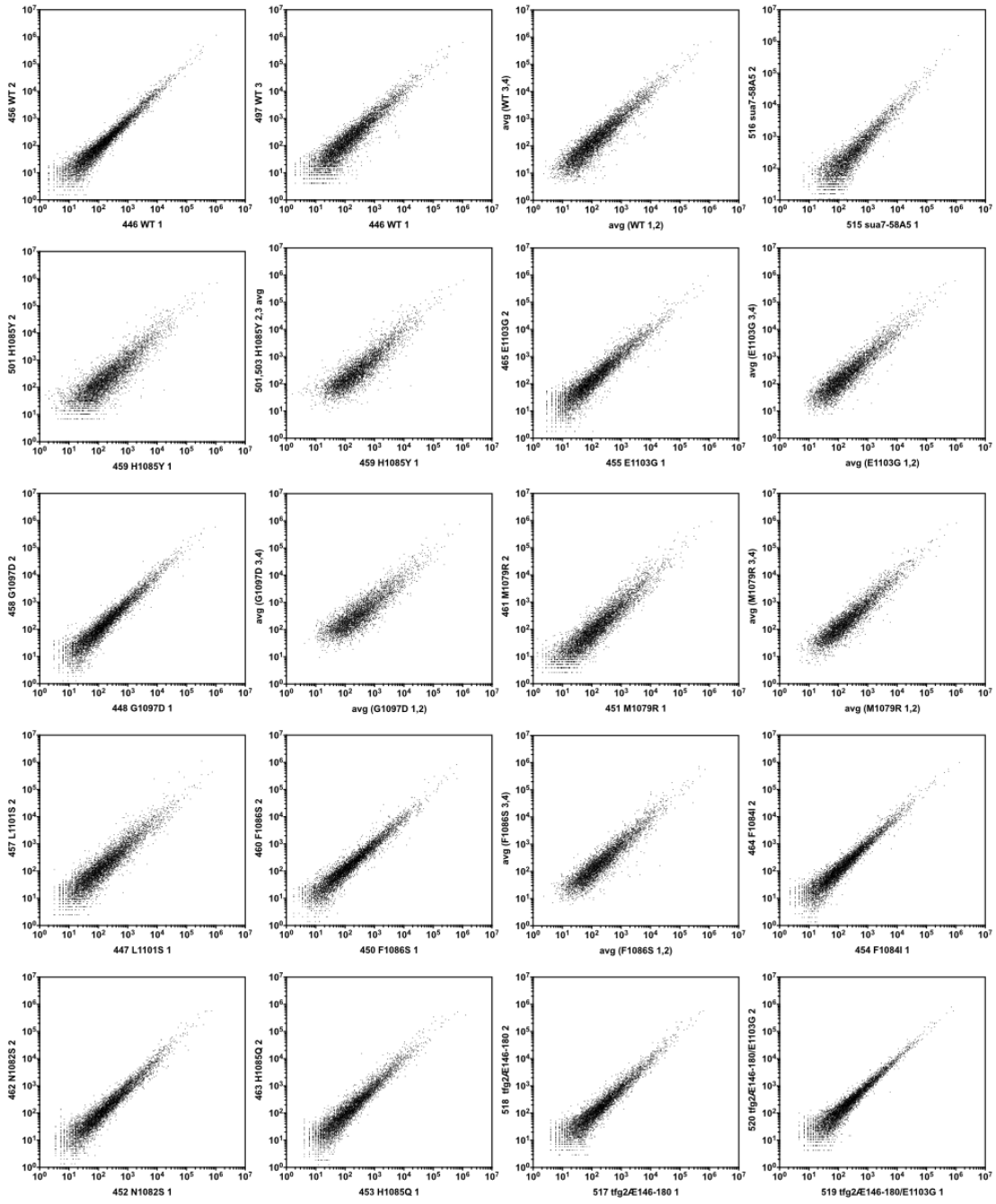


Figure 4-5. XY graphs for accumulative reads per promoter window (expression) are shown for replicate libraries.

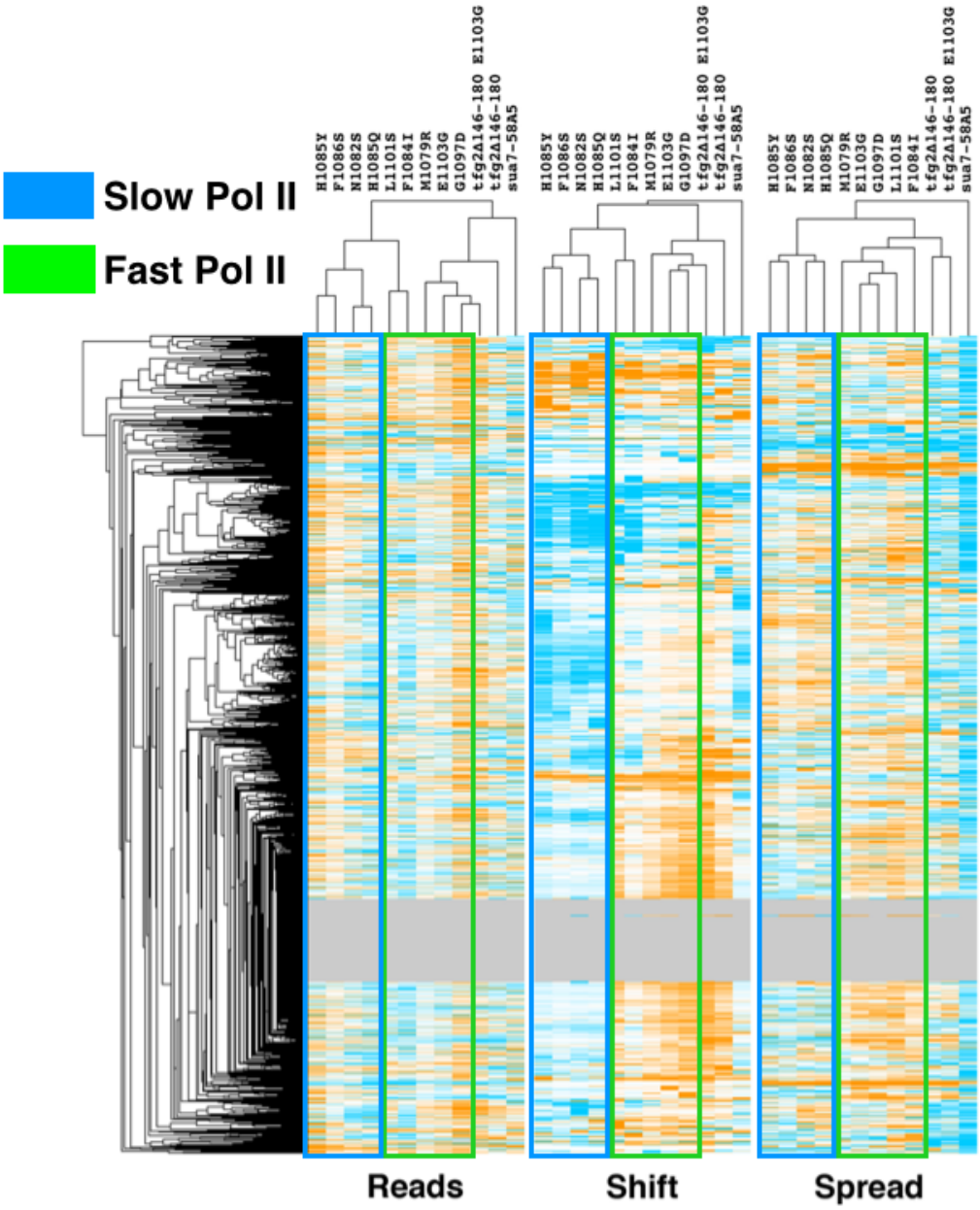
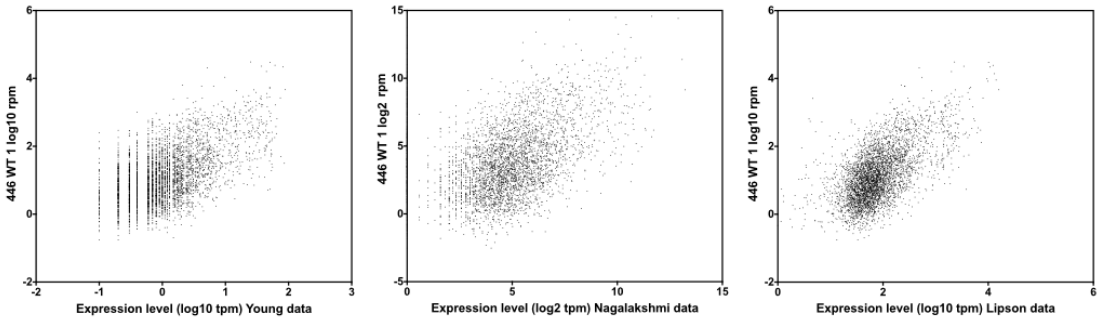


Figure 4-6. TSS characteristic alterations cluster two classes of Pol II activity/GTF mutants separately.
Accumulative reads (reads for each library was scaled to library 446), TSS shift and spread numbers for each promoter were loaded into TreeView, followed by vertical hierarchical clustering by Cluster 3.0. The promoter order generated by clustering was kept, numbers in each library were then horizontally clustered independently for each characteristics. Orange indicates increase relative to WT, cyan indicates decrease. Slow Pol II mutants are marked by blue box; fast mutant by green. Note: We are currently in process of repeating this clustering with quantile normalized rpm (see below).

To test if replicates of mutant or WT strains correlated within replicates, we assessed base pair correlations between replicate experiments for read levels at individual genome positions. We took the natural logarithm of uniquely mapped TSS counts at each nucleotide and calculated the Pearson correlation between libraries (**Figure 4-4**). Replicates showed high correlation in our nucleotide based comparison when filtering for genome positions with differing minimum levels of read counts (2 or 9 read count minimums). All TSSs assigned in promoter windows hereafter are strand-specific. Furthermore, when we examined overall reads in promoter windows in replicate libraries we observed even higher correlations between replicates (**Figure 4-5**). For subsequent analyses, we combined uniquely mapped reads within the promoter windows from replicate libraries and performed downstream analyses with merged libraries. Read numbers for each genotype after library merging are listed in **Table 4-3**. We measured characteristics of TSS utilization as follows: distance to core promoter element (CPE) was calculated by subtracting median TSS position from position of proposed TATA/TATA-like element. TSS "spread" was defined as width of dispersed TSSs distribution and was calculated by subtracting 10^{th} percentile read position from 90^{th} percentile read position in the promoter window in a strand specific manner. Gene expression was calculated by sum of all TSS reads in a promoter window (**Figure 4-1**). Our previous analyses indicated that gene expression profiling and genetic interaction profiling could finely separate Pol II mutants into classes both by type of mutant (fast/slow) and by the magnitude of their biochemical defects (Chapter II, (Braberg et al., 2013)). We found that changes in either TSS positions ("Shift"), width of TSS

distribution ("Spread"), or total reads in mutant libraries relative to WT were all able to distinguish Pol II fast and slow mutants from each other (**Figure 4-6**). First, slow mutants show downstream TSS shifts for a majority of promoters while fast mutants overall show upstream TSS shifts (in this calculation, mutants that shift TSS downstream have negative shift; mutants that shift TSS upstream have positive shift). Second, most mutants showed increased TSS spreads at majority of promoters. Mutants showed greater usage at upstream or downstream positions depending on class, while still using original TSSs to some extent, therefore by our definition of "spread", they are expected to increase the spread. GTF mutants are clustered on the right side of clustered heatmap, however visual observation indicates *tfg2*Δ*146-180* (subunit of TFIIF) and *sua7-58A5* (TFIIB) show TSS alterations similar to the two separate classes of Pol II activity mutant. *sua7-58A5* (TFIIB) shows strong similarity to Pol II slow mutants and *tfg2*Δ*146-180* (TFIIF single mutant) and *tfg2*Δ*146-180*/E1103G show strong similarities with Pol II fast mutants (**Figure 4-6**). These results are consistent with their alteration of TSS distributions at model genes tested previously (Chapter II, III, (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012)). Changes conferred by Pol II activity/GTF mutants on global scale matched those we had previously observed for *ADH1* (as mentioned in Chapter III). Our estimation of gene expression, determinations of which can be noisy when examined by different experimental platforms and if conditions or strain backgrounds are not identical, has reasonable if slightly low correlations with previously published gene expression estimates (**Figure 4-7**) (Holstege et al., 1998; Lipson et al., 2009; Nagalakshmi et al., 2008). We note that it has been suggested that

normalization methods that calculate fraction in total such as calculation of rpm and tpm are not ideal to estimate transcript level using relatively old sequencing data (Liu et al., 2012; Rao et al., 2008; Rapaport et al., 2013). When we examined the relationships between libraries by our position-based correlation analysis, we were also able to cluster fast Pol II/TFIIF mutants and slow Pol II/TFIIB mutants into distinct groups (**Figure 4-8**). In conclusion, we are confident with the quality of our TSS sequencing and the calculations for TSS characteristic measurements and their changes in mutants. We showed that Pol II activity/GTF mutants confer genome wide directional shifts in TSS distribution, see below for more comparisons and discussion.



Expression dataset for comparison	Sequencing method	Data point pair #	Pearson R
Lipson	single-molecule sequencing digital gene on poly-Aexpression (smsDGE)	4252	0.6072
Nagalakshmi	RNA-seq	4252	0.3930
Young	high-density oligonucleotide arrays	4024	0.3250

Figure 4-7. Expression correlation with three published datasets.
Sequencing methods, data point pair numbers used for comparison with each data set comparison and Pearson R correlation with each dataset are shown below.

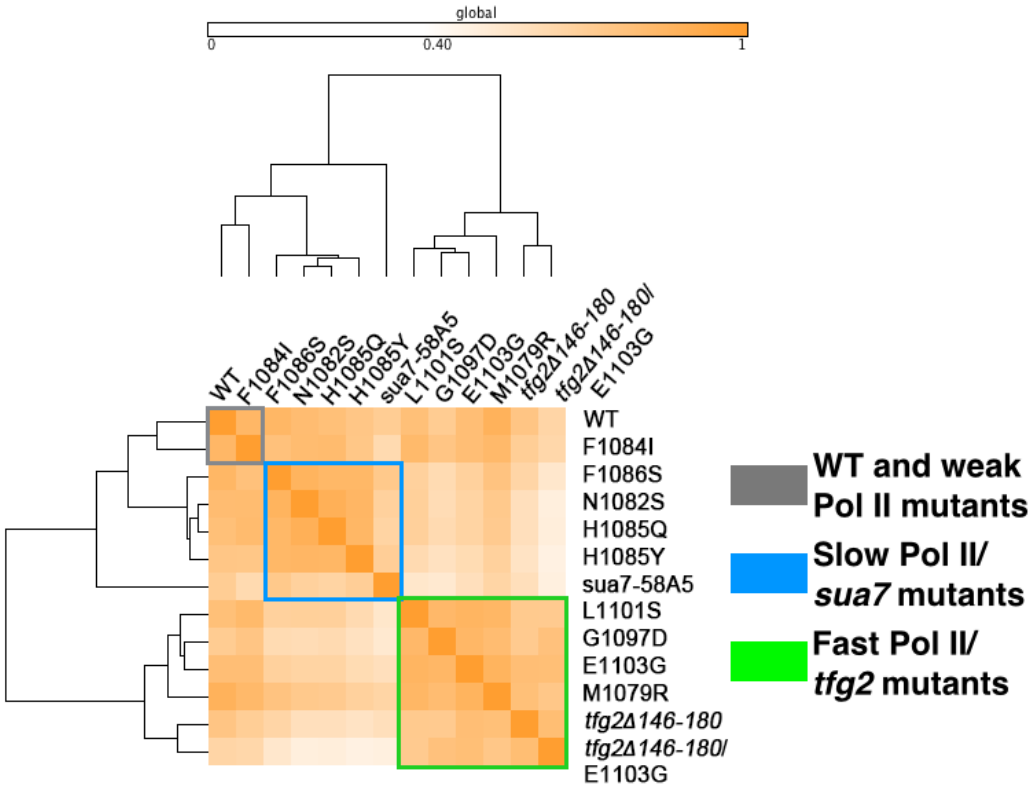


Figure 4-8. Position based correlation matrix of merged reads per WT/mutant. Position based correlations are calculated as in Figure 4-4. Correlation was calculated for data points with read>3 counts for all possible data pairs. Matrix was clustered both vertically and horizontally in GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/index.html). WT and a weak Pol II mutant that has exhibited a mild TSS phenotype are marked by grey box; slow Pol II/*sua7* mutants are marked by blue box; fast Pol II/*tfg2* mutants are marked by green box.

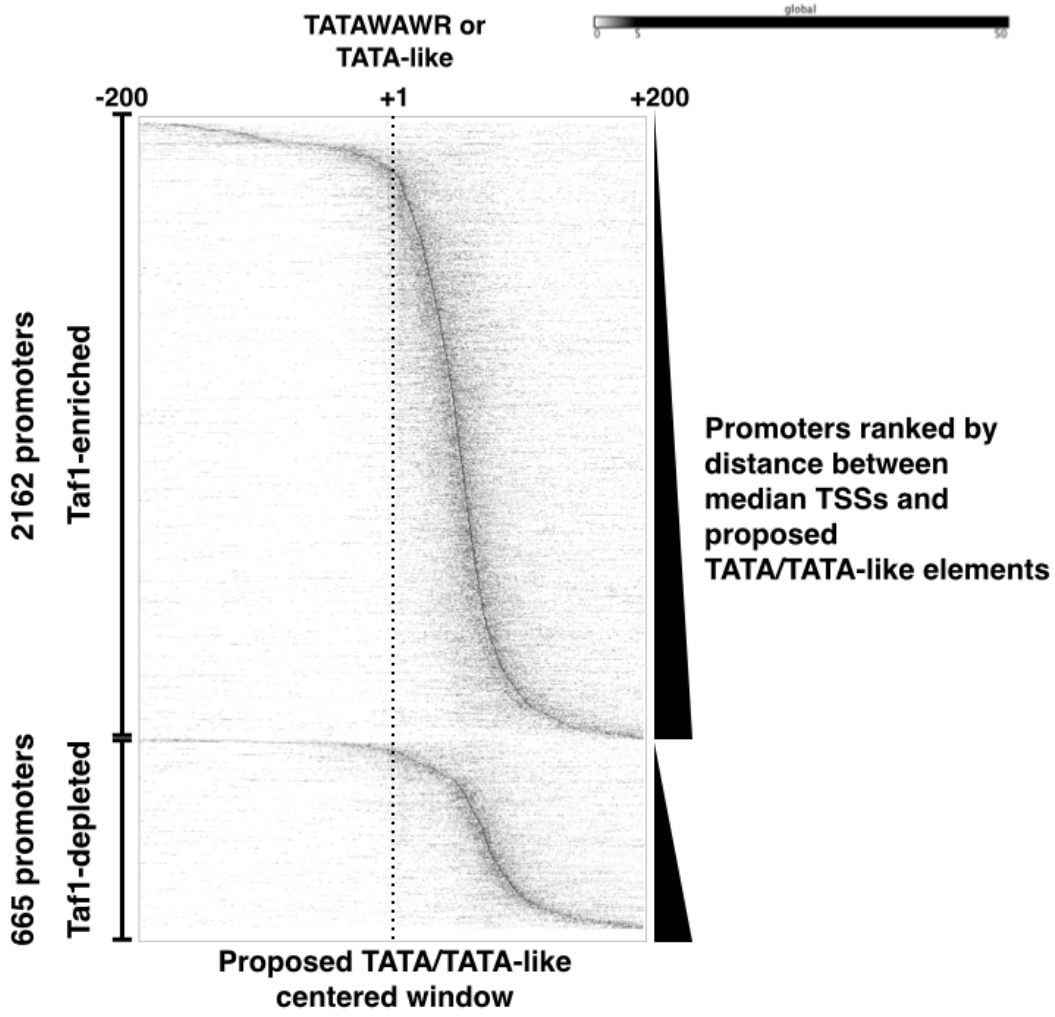


Figure 4-9. TSS mapping in proposed TATA-like/TATA element centered windows. Promoter centered windows were designated by assaying 200 nt on both sides from proposed TATA/TATA-like elements. Promoters above an expression cutoff (>10 rpm) are sorted into Taf1-enriched and Taf1-depleted classes, then ranked by distances between median TSSs and proposed TATA/TATA-like elements.

First, we wished to examine relationship between TSSs mapped by us and the proposed TATA/TATA-like elements (Rhee and Pugh, 2012) to assess the positional relationships between them. TSSs are expected to be mapped downstream of core promoter elements in directional promoters, and by comparing relative positions of core promoter element and TSSs, we are able to assess whether a core promoter element is positioned appropriately to drive the TSSs. **Figure 4-9** shows positioning of TSSs (showing normalized densities in each promoter) relative to proposed TATA and TATA-like element (Rhee and Pugh, 2012) in Taf1-enriched and Taf1-depleted promoters above 10 read per million (rpm) expression cutoff (~350 reads in WT promoter window, leaves 3070 of 6044 promoters, also see figure below for cutoff) in each proposed core promoter element centered window.

We make the following observations. First, we note our findings of a considerable number of misannotations in proposed TATA/TATA-like elements (~12% TATA, 2.6% TATA-like element, **Table 4-1**). The misannotated TATA elements mainly stem from the predicted TATA element being in the incorrect orientation for the proposed direction of the promoter/gene. Second, there are likely differences in promoter element prediction success between two promoter classes (Taf1-enriched and Taf1-depleted). The majority of Taf1-depleted (majority are TATA-containing) promoters showed TSSs mapped 40−120 nt from the TATA element, consistent with previous observations of functional TATA elements and their distances upstream of TSSs at specific promoters by visual inspection. This is consistent with the identified TATA elements functioning as a core promoter element to drive transcription (Rhee and Pugh,

2012). Only a small fraction of promoters show TSSs mapped at exceptional distances from predicted TATA elements (visual inspection). The Taf1-enriched (mostly TATA-less) promoter class shows a greater number of promoters where TSSs are in positions unlikely to be driven by the proposed TATA-like element (visual inspection). These include: first, TSS that are upstream of the proposed TATA-like promoter element (<0 from TATA-like element); second, TSSs that are seemingly too close to support any known mechanism of initiation (0− +30 nt from TATA-like element); third, TSSs that are seemingly too far downstream (>+120 nt from TATA-like element) (**Figure 4-9**). We suspect that these TSSs (roughly estimated to be of~1/3 Taf1-enriched promoters in this figure, visual inspection) are not likely to be driven by the proposed TATA-like elements, and led us to ask if these TATA-less promoters have other promoter elements (see experiments below). Due to these observations, we preferred to classify promoters into Taf1-enriched or Taf1-depleted classes in our analyses. As shown in **Table 4-1**, the majority of TATA-less promoters are Taf1-enriched while consensus TATA-containing promoters are mostly Taf1-depleted thus two types of classifications render similar sets of promoter groups.

Next, we wished to address TSSs and other initiation properties in a TSS aligned window. This allows examining properties relative to TSS positions without relying on correct determination or estimation of core promoter element positions, and is another powerful way to examine TSS selection characteristics. We took median the TSS of each promoter in our WT merged library (**Table 4-3**) and expanded 250 nt upstream and 150 nt downstream to designate “TSS aligned windows” (401 nt range). **Figure 4-10** shows

positioning of TSSs normalized density (for each promoter) in TSS aligned window. Very low reads are detected at far distances from median TSS (<−100 nt from median TSS; >+100 nt from median TSS), and higher expressed promoters showed more concentrated TSSs in spreads less than 50 nt around the position of the median TSS (**Figure 4-10**).

ChIP-exo sequencing to map GTFs localizations in genome wide

We hypothesized that Pol II activity/GTF mutants alter scanning, thus localization of GTFs may be altered in our mutants. Furthermore, because a number of TATA-like promoter elements in Taf1-enriched/TATA-less promoters annotated by Pugh group either have misannotations and/or deviate from expected positions relative to our TSS (**Figure 4-9**) as mentioned above, we needed a better estimation for core promoter (proxy of PIC assembly) especially for Taf1-enriched promoters for our analyses below, which assess promoter architecture and its relation to TSS usage. We determined to map GTFs in WT and Pol II activity mutant strains (same strain background as in our SOLiD TSS sequencing) and process ChIP-exo sequencing reads differently from the investigation predicted positions of TATA/TATA-like elements (Rhee and Pugh, 2012). In their published analyses, the Pugh group processed reads on both strands separately and generated peak pairs by peak calling based on reads in each strand, thus promoter prediction may have been subject to read noise in the data. In our processing of ChIP-exo sequencing, we took a more conservative approach and looked at raw read distributions, having determined which strand (relative to the strand promoters are mapped to) the mapped Watson and Crick strand reads were on.

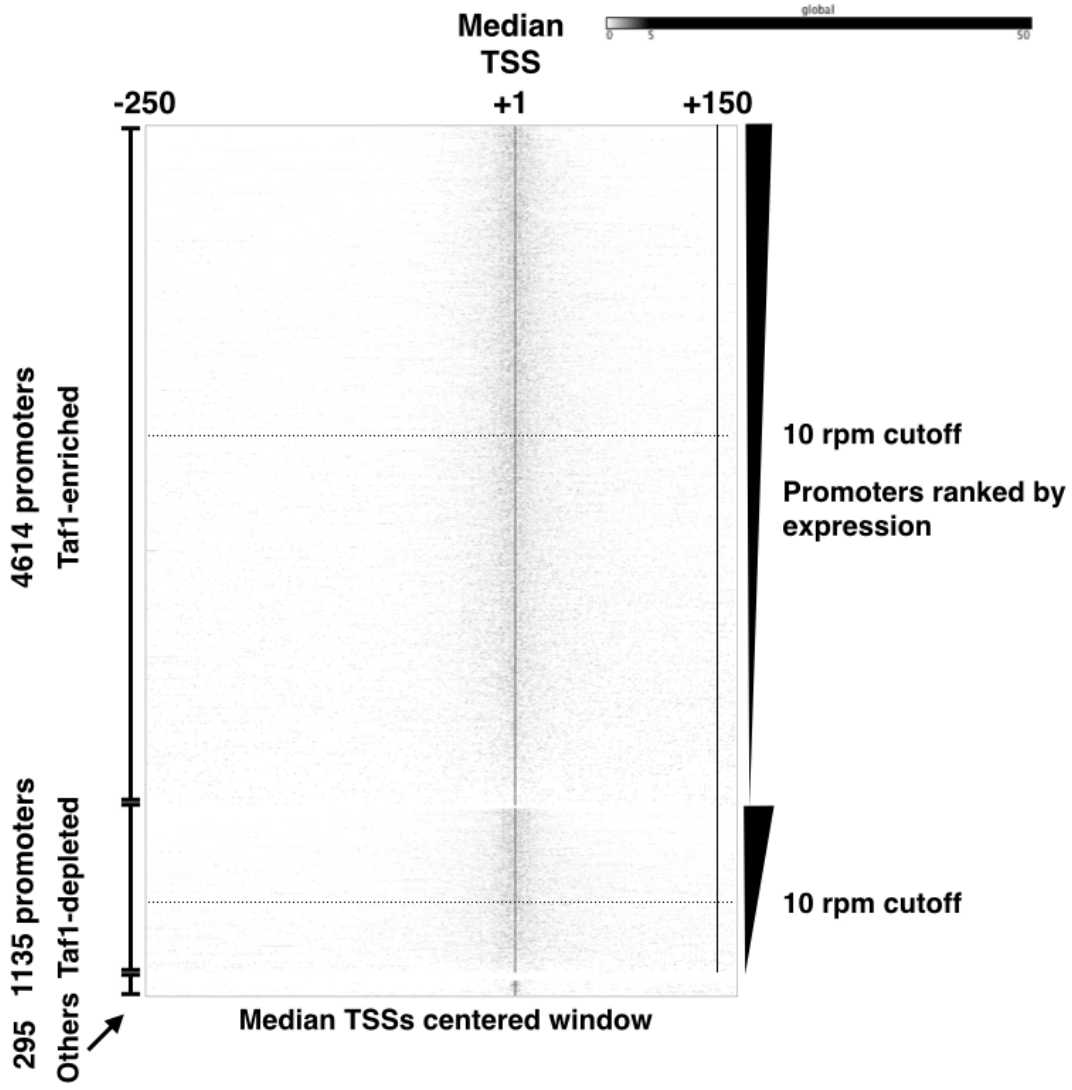


Figure 4-10 TSS mapping in median TSS aligned windows.
Median TSS (in WT merged reads, see Table 4-3) position for each promoter and 250 nt upstream, 150 nt downstream are used to designate median TSS aligned promoter windows. TSS densities for each promoter are normalized to total of 100%. All promoters in the study are shown, and promoters are sorted into Taf1-enriched and Taf1-depleted classes, and ranked by expression. Promoters that pass 10 rpm expression filter are separated by dotted line from promoters less than 10 rpm expression in each promoter class.

Our processing is as follows. In ChIP-exo sequencing, the 5′ ends of reads on either strand indicate the position in the genome where a protein-DNA crosslink protects a DNA fragment from exonucleolytic digestion by an exonuclease (see **Methods**). Strand conscious analyses of 5′ ends of reads detect ends on both sides of a crosslinked factor. Ends on TOP strand reads (reads of a promoter in the direction of transcription, thus TOP is relative to a promoter and may map to either Watson (forward)/Crick (reverse) strains of the genome) correspond to the upstream edge of a bound target protein's footprint, while ends on BOTTOM strand correspond to downstream edge of a bound target protein's footprint. We preferred to process ChIP-exo reads in this way, which is somewhat similar to process of TSS sequencing to assess reads in a less read coverage biased manner and are able to include all reads for each promoter (in promoter window) instead of only examining peak pairs.

We mapped two GTF subunits Ssl2 (a ATPase/helicase subunit of GTF TFIIH) and Sua7 (encodes TFIIB) in our WT and three Pol II mutant strains in collaboration with Pugh lab (PSU). I have not obtained a second repeat of sequencing that has satisfying read coverage or signal/noise ratio after two trials, therefore the ChIP-exo data used in this document is from one round of experimentation from our strains. We are currently waiting for a second repeat experiment with satisfactory data quality. For the purposes of our analysis of WT strains, we are in process of compairing our dataset to the published datasets in (Rhee and Pugh, 2012). See **Table 4-4** for reads stats for ChIP-exo sequencing, see **Methods** for more details about experiment.

Table 4-4 ChIP-exo libraries reads

Target	Mutation	Index Count	Unique Read Count	Unique Read Percent
Sua7	WT	3,988,015	2,899,922	72.70%
Ssl2	WT	13,286,752	9,721,213	73.20%
Sua7	*rpb1*-M1079R	4,163,096	2,801,559	67.30%
Ssl2	*rpb1*-M1079R	13,261,484	9,579,314	72.20%
Sua7	*rpb1*-G1097D	4,439,799	3,128,228	70.40%
Ssl2	*rpb1*-G1097D	20,312,457	14,378,893	70.80%
Sua7	*rpb1*-E1103G	1,256,252	920,032	73.20%
Ssl2	*rpb1*-E1103G	7,221,180	5,388,435	74.60%
Sua7	*rpb1*-H1085Y	5,150,891	3,703,774	71.90%
Ssl2	*rpb1*-H1085Y	6,104,810	4,554,160	74.60%

Figure 4-11 shows Ssl2 (subunit of TFIIH) mapped in median TSS window, promoters are ranked by distance between median TSS to median Ssl2 ChIP-exo reads on TOP strand. **Figure 4-12** shows Sua7 (TFIIB) mapped in median TSS window, promoters are ranked by distance between median TSS to median Ssl2 ChIP-exo reads on TOP strand. Majority of ChIP-exo signals localize upstream of TSSs suggesting PIC assembly is detected for all promoters upstream of where TSSs are observed, supportive of directional scanning model (**Figure 4-12**). This observation is also consistent with

published ChIP-exo for GTF subunits, thus we take our single ChIP-exo data as confirmatory. Ssl2 and Sua7 (both subunits of PIC) localization is very similar as expected. I use the position of the median Ssl2-TAP WT ChIP-exo signal in promoter windows as an estimate of PIC assembly position for Taf11-enriched promoters in our following analyses.

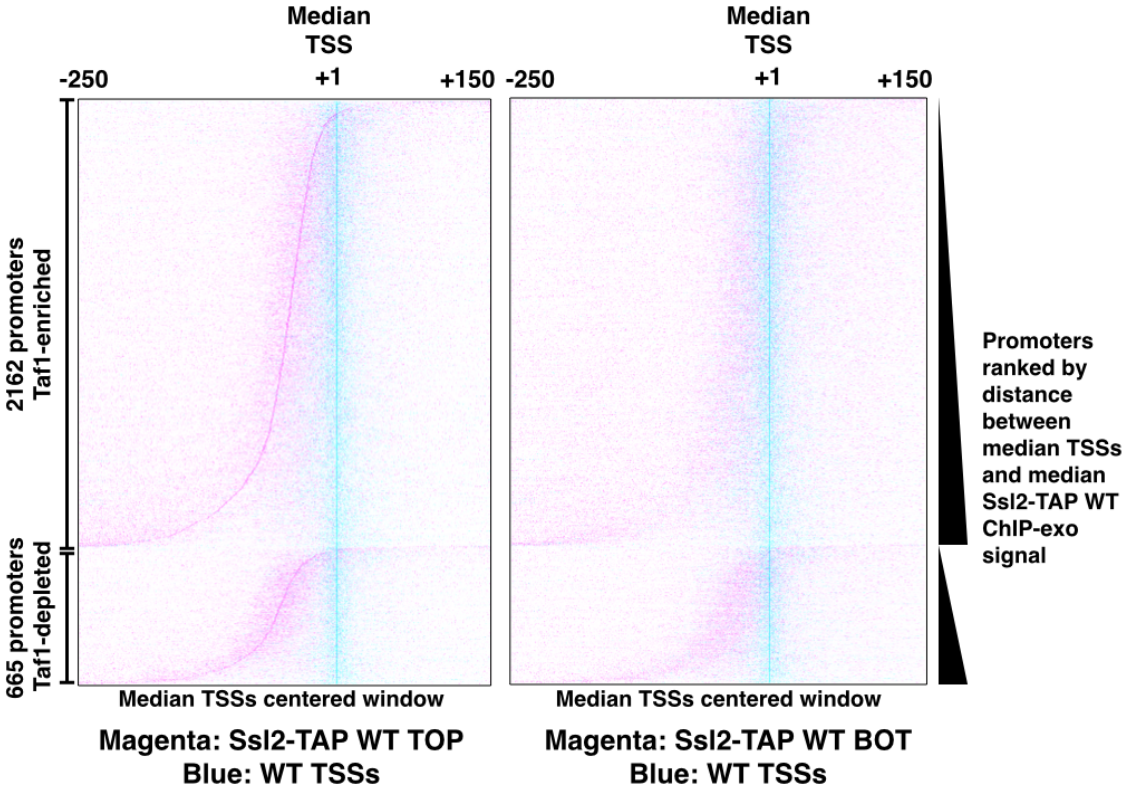


Figure 4-11. Ssl2 ChIP-exo mapping in median TSS aligned windows.
Normalized densities for Ssl2 (a subunit of TFIIH) ChIP-exo signals on TOP strand (left panel) and BOTTOM strand (right panel) are shown in magenta. Normalized TSS densities are shown in blue. Promoters are sorted by promoter classes and ranked by distance between median TSS and median Ssl2 ChIP-exo TOP signal.

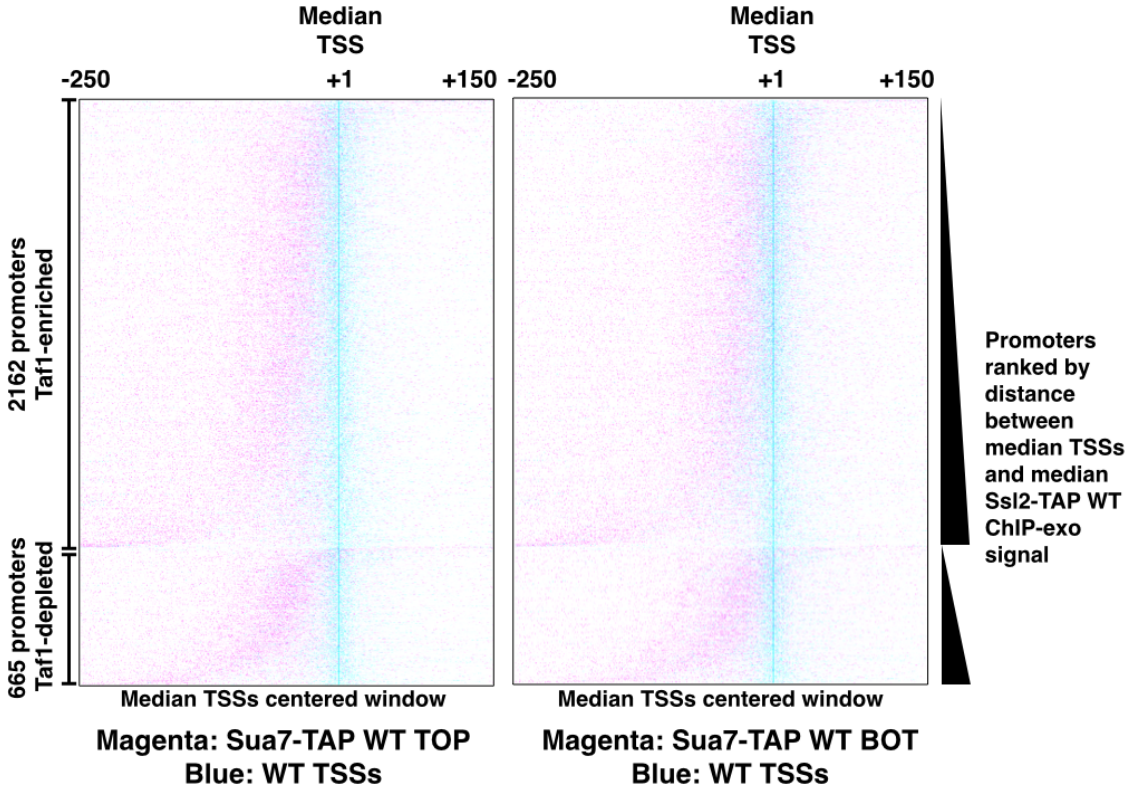


Figure 4-12. Sua7 ChIP-exo mapping in median TSS aligned window.
Same as Figure 4-11, but for Sua7 (TFIIB), ChIP-exo signals on TOP strand (left panel) and BOTTOM strand (right panel).

Test of an existing model for TSS utilization in different promoter classes

How and if TSS selection differs for different promoter classes is unclear. An existing model proposed by the Pugh group for TSS selection/utilization in different promoter classes is that Taf1-enriched/TATA-less class assembles PICs (Pre-Initiation Complex) cooperatively with the +1 nucleosome with the +1 nucleosome constraining the distance for Pol II scanning. This constraint leads to TSSs closer to the position of the predicted core promoter. Conversely, Taf1-depleted/TATA-containing promoters assemble their PICs competitively with promoter nucleosomes, allowing Pol II to scan in a wider range of promoter region without physical constraint from the nucleosome because the nucleosome will be absent when the PIC assembles (Rhee and Pugh, 2012) (**Figure 4-13**). Expectations of this model are: First, distances between TSS and core promoter elements in Taf1-enriched/TATA-less class (**Figure 4-1**) should be shorter than those for Taf1-depleted/TATA-containing class. Second, "spread" of TSS (**Figure 4-1**), the width of TSS distribution aka how dispersed TSS are) in Taf1-depleted/TATA-containing promoters might be larger because scanning is not restricted by a nucleosomal barrier. Third, +1 nucleosome occupancies in Taf1-depleted/TATA-containing promoters class should be critical determinants for expression level of those promoters.

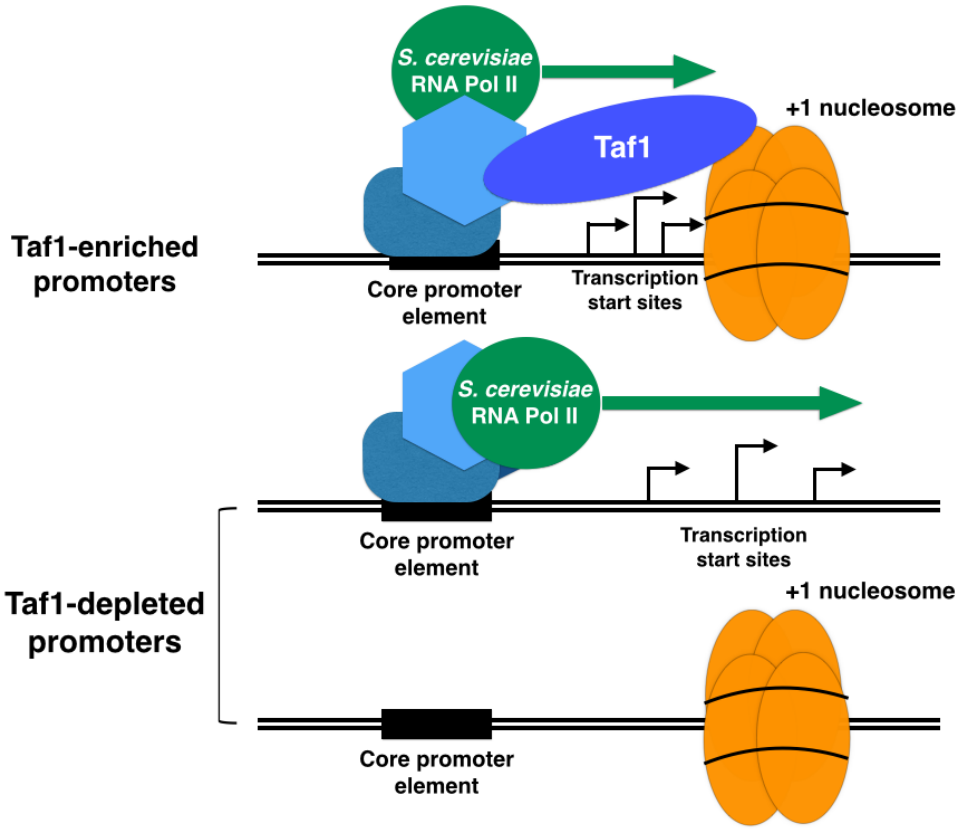


Figure 4-13. Current model for TSS selection in different promoter classes.

First, we investigated TSS characteristics for the two main *S. cerevisiae* promoter classes in **Figure 4-15**. Box plots in this chapter are all Tukey box and whisker plots (**Figure 4-14,** see figure legend for explanation). Promoters with highest expression levels show narrowest TSS spreads in both promoter classes, and TSSs are in general more dispersed in lowly expressed promoters. The trend seems more uniform and better correlated with expression level for Taf1-enriched promoters (**Figure 4-14**). Taf1-depleted promoters are largely enriched in highly expressed class, while Taf1-enriched promoters are evenly distributed in all expression levels. Median distances to presumptive core promoters in all expression deciles are ~50 nt and show increase of the IQR (inter-quartile range) by expression decile in the Taf1-enriched class. This and the observation that Ssl2 cumulative reads (a subunit of GTF TFIIH) are higher in higher expressed promoters are consistent with a model that highly expressed promoters have more stereotypical and/or higher occupancies of PIC assembly (**Figure 4-14**).

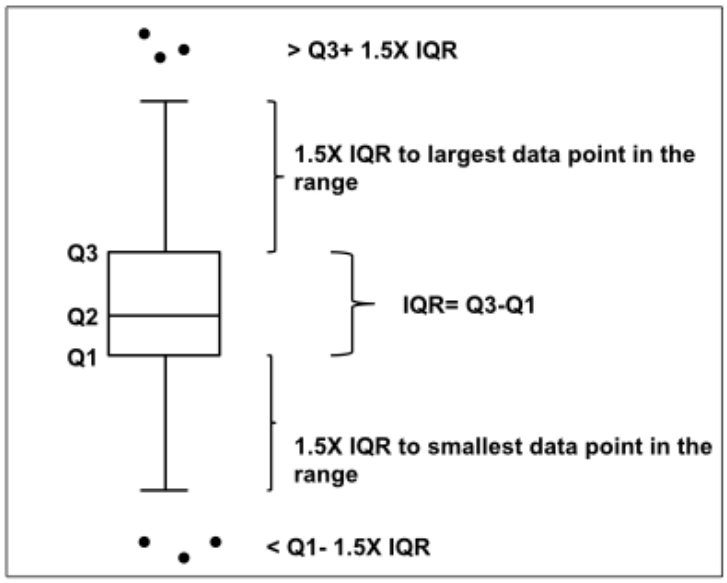


Figure 4-14. Tukey box and whisker plot.
All box plots in this chapter are of this type. The box extends from 25th percentile number (Q1) to 75th percentile number (Q3), interquartile range (IQR) is the difference between Q1 and Q3. Median (50th percentile, Q2) is marked within the box. Calculate a value Q3+1.5×IQR, if this value is greater than or equal to the largest data point in the data set, upper whisker extends to largest data point in the set; if this value is less than the largest data point, upper whisker extends to the largest data point less than this value, all data points outside whisker are termed outliers, and individual data points are shown for outliers. Lower whisker is same concept, the value is Q1−1.5×IQR.

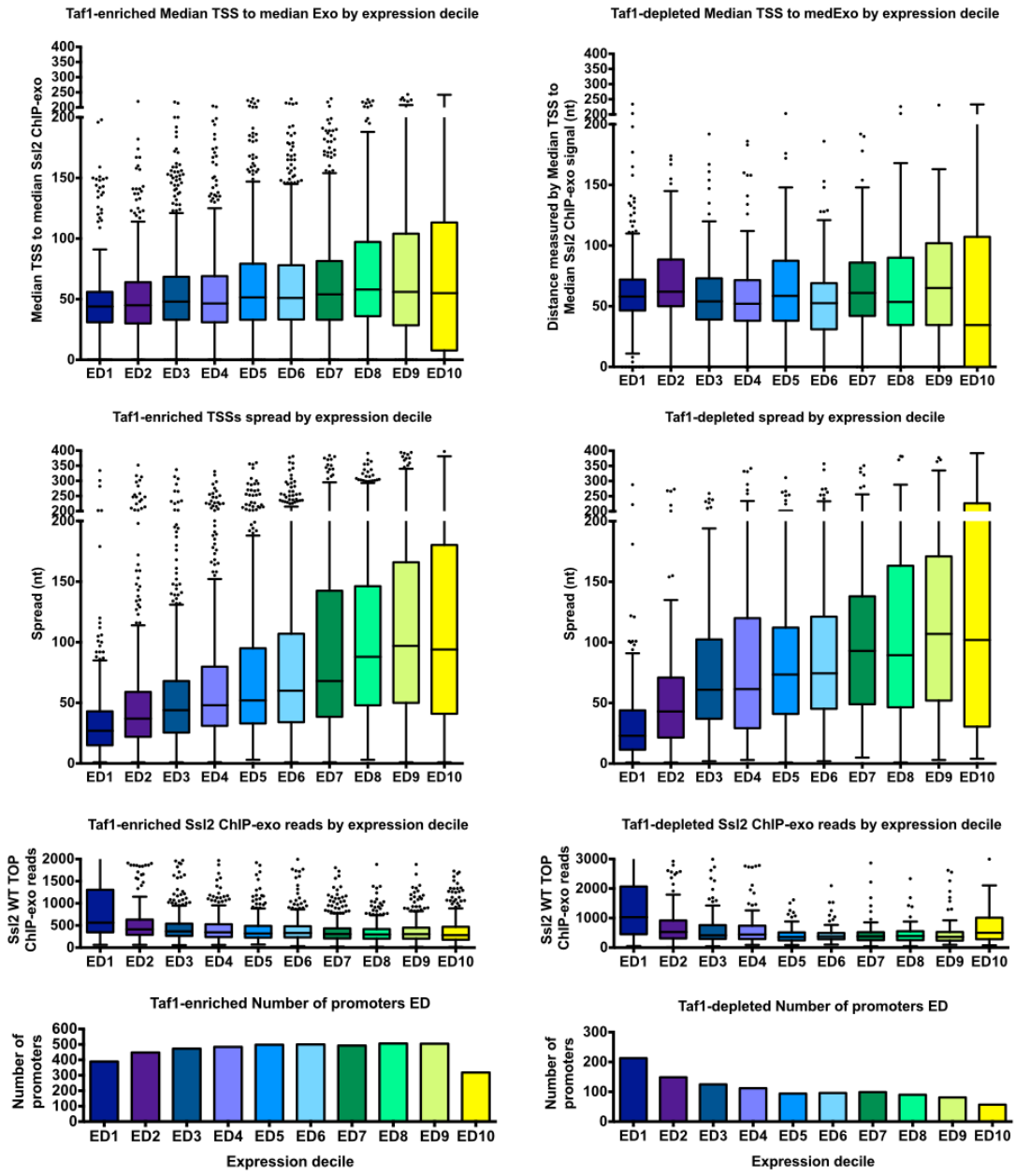


Figure 4-15. TSS characteristics of promoters binned by expression decile.
Left panel: Taf1-enriched promoters; Right panel: Taf1-depleted promoters.

We tested this model with our mapped TSSs within promoter windows. We sorted promoters into four classes to keep promoter clustering consistent with original publication where the model was proposed (Rhee and Pugh, 2012). First we tested if the distances between TSS and core promoter elements were shorter in Taf1-enriched/TATA-less promoter class as predicted by the model. One point worth mentioning is that the Pugh group used distance between a single annotated TSS for each promoter and proposed TATA-like element as distance between core promoter element and TSS for the TATA-less promoters in their analysis. The criteria for which TSSs were picked was not clear (TSS data was from (David et al., 2006)). Therefore, we reexamined this issue using our high-resolution TSS data and a more conservative landmark for core promoter position –median ChIP-exo read position in place of predicted core promoter element. Using our TSSs and previously predicted CPEs, we do indeed see significantly shorter CPE-TSS distances for Taf1-enriched/TATA-less promoters (**Figure 4-16** A). However, when median ChIP-exo signal for Ssl2 (a subunit of GTF TFIIH) is used, the differences in distances between core promoter element and TSSs in promoter classes are somewhat diminished and the difference becomes not as strong (**Figure 4-16** C). We chose promoters that have high ChIP-exo signals within the promoter window so that we might estimate presumptive PIC position and the core promoter with higher confidence. **Figure 4-16** D promoters were filtered based on top 50% ChIP-exo signals, and distances between core promoter elements and TSSs in two classes were not significantly different.

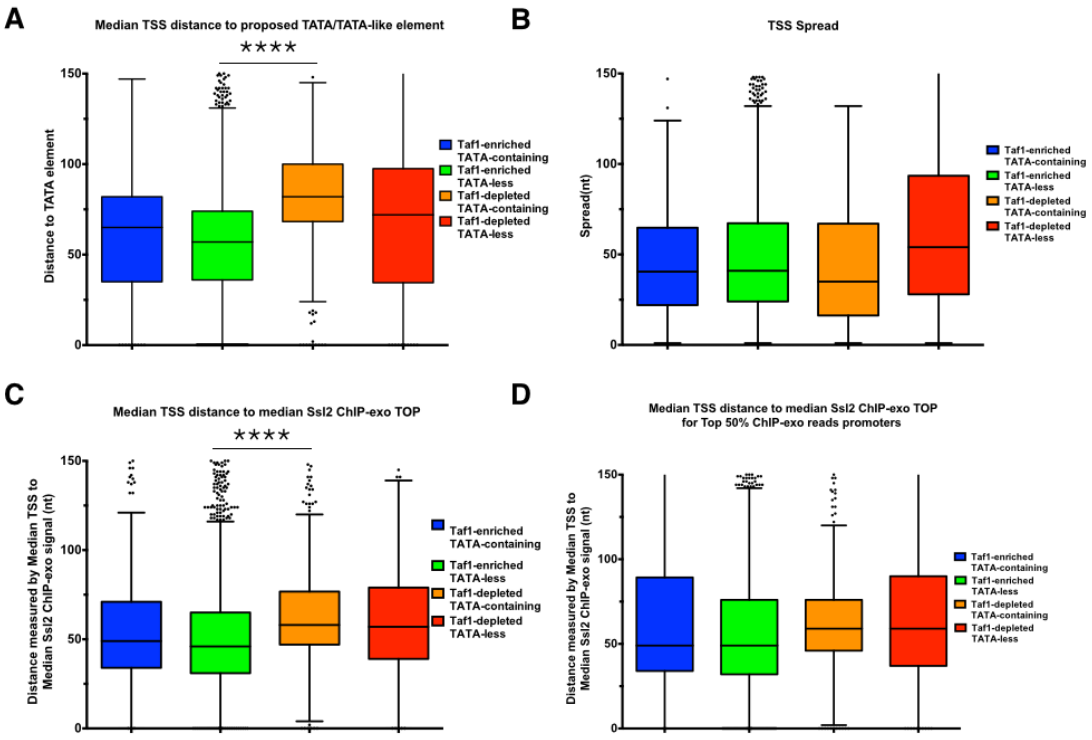


Figure 4-16. Testing current model for TSS selection (Figure 4-13) with our data. Promoters are sorted into four classes, various distance comparisons are shown: **A.** median TSS distance to proposed TATA/TATA-like elements; **C.** median TSS distance to median Ssl2 ChIP-exo signals; **D.** median TSS distance to median Ssl2 ChIP-exo signals for promoters in the top 50% Ssl2 ChIP-exo reads. TSS spread comparison is shown in **B.**

We next tested if the TSS distribution widths ("spreads") for Taf1-depleted/TATA-containing promoters were larger than Taf1-enriched/TATA-less promoters as predicted by the simplest version of the model (**Figure 4-16** C). No significant difference was detected, and in fact, a considerable fraction of Taf1-depleted/TATA-containing promoters showed narrower spreads. One caveat worth noting for tests of first two model predictions is that the analysis only considers the TSS characteristics regardless of promoter architecture. For example, the evolved promoter sequence context is an important determinant for regions that allow TSS selection because *S. cerevisiae* initiator motif is $Y_{-1}\underline{R}_{+1}$. Differences in promoter architecture and sequence context could be the main reason for observation of narrower distances between core promoters and TSSs for Taf1-enriched promoters, and not restriction of scanning due to nucleosomes. Similarly, lack of significant difference in TSS spreads does not necessarily rule out restricted scanning.

Pol II catalytic activity mutants confer shifts in TSS utilization distribution genome wide

We next asked if the shifts in TSS usage distribution conferred by Pol II activity/ GTF mutants at model genes were genome-wide phenomena. Most primer extension experiments for directional TSS shifting mutants tested previously were mostly at TATA-containing model genes, supporting a scanning mechanism at TATA-containing promoters, leaving the question open for non-TATA promoters. As non-TATA promoters have different architectures and characteristics (discussed above), it is possible that these promoters may respond differently to altered Pol II activity or initiate transcription with different features. Our genome wide TSS mapping shows that both Pol

II activity mutants and GTF mutants confer consistent polar effects on TSS distribution at most yeast promoters, regardless of class, suggesting that promoter scanning is likely operating universally in yeast (**Figure 4-17**).

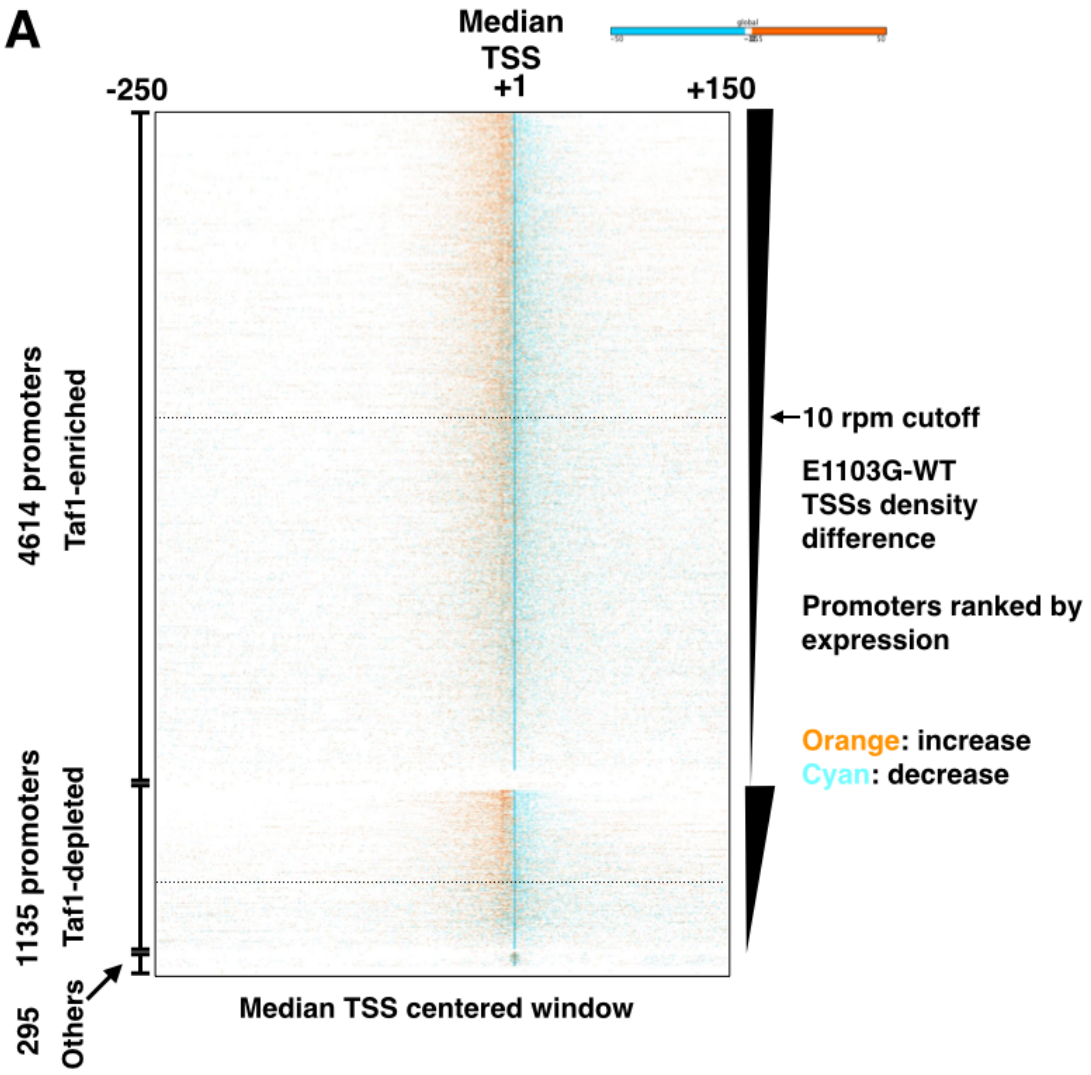


Figure 4-17. Polar shifts in TSS usage distribution in Pol II activity mutants. Promoters are sorted into Taf1-enriched and Taf1-depleted classes and ranked by expression. TSS normalized density differences are shown in median TSS aligned window for **A.** A fast Pol II mutant *rpb1* E1103G and **B.** A slow Pol II mutant *rpb1* H1085Y.

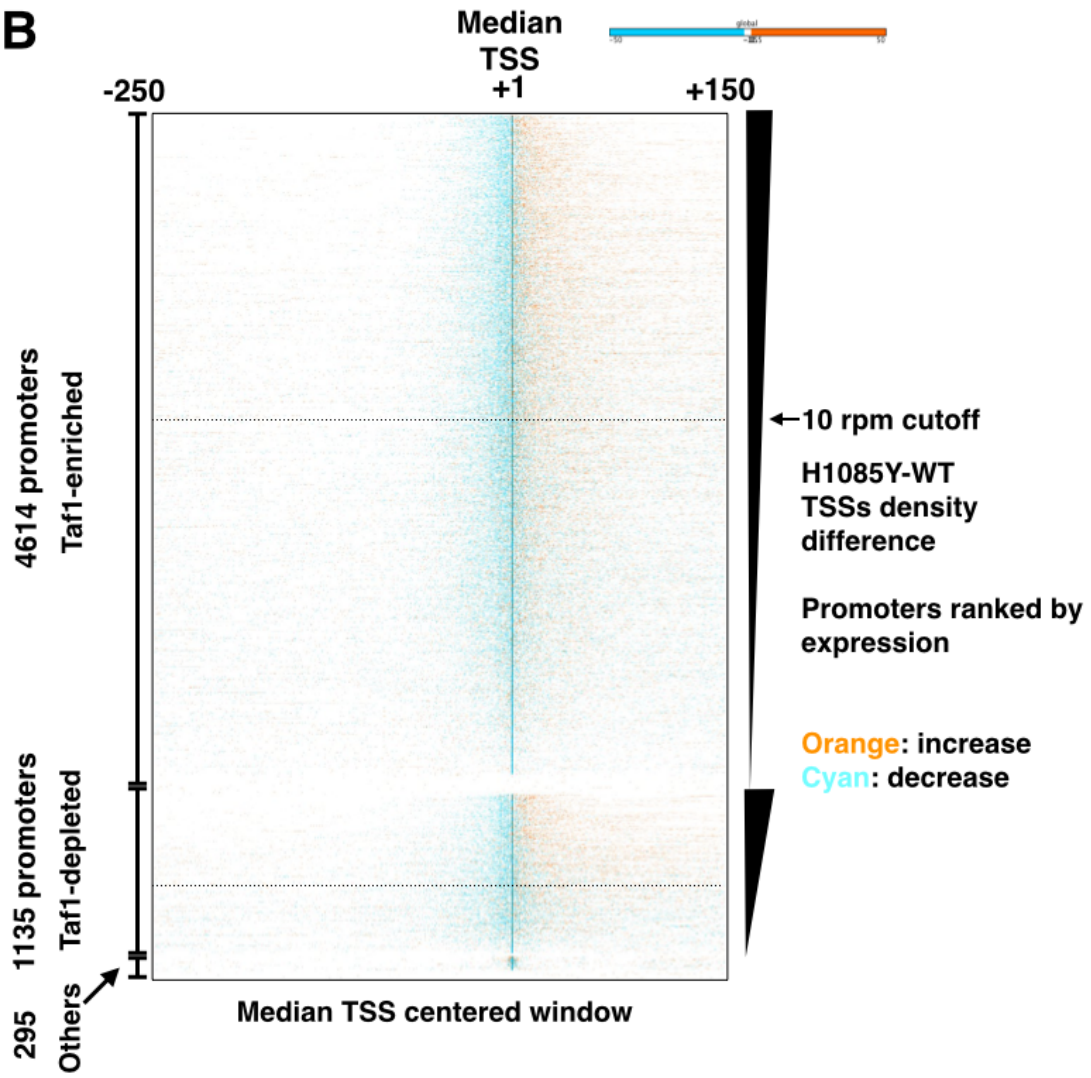


Figure 4-17 Continued.

The magnitude of shifts conferred by Pol II activity/GTF mutants at the majority of promoters conferred by Pol II activity/GTF mutants correlates well with the deviation of their measured or genetically estimated catalytic activities compared to WT, and is consistent with TSS shifts measured by primer extension at *ADH1* previously (**Figure 4-17**, **Figure 4-18**). **Figure 4-17** shows genome-wide TSS shifts conferred by a representative Pol II fast mutant E1103G (A) and representative Pol II slow mutant H1085Y (B). **Figure 4-18** shows TSS shifts in all tested mutant strains for all promoters having expression above an arbitrary expression cutoff, separated by promoter class. TSS shifts in the mutant strains were measured at each promoter by subtracting distance between the proposed TATA/TATA-like element from the median TSS measured in WT from the median measured for each mutant (best estimation of core promoter element is not required here since it is only used as a starting point and does not affect the calculation of TSS shift). Therefore, a downstream shifting mutant will give a positive TSS shift in this calculation because the distance from the proposed core promoter element will be longer, while an upstream shifting mutant confers a negative TSS shift at a promoter. The majority of promoters show consistent direction and magnitude of shifts as those observed at *ADH1* (Chapter II, III, (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012))(**Figure 4-18**).

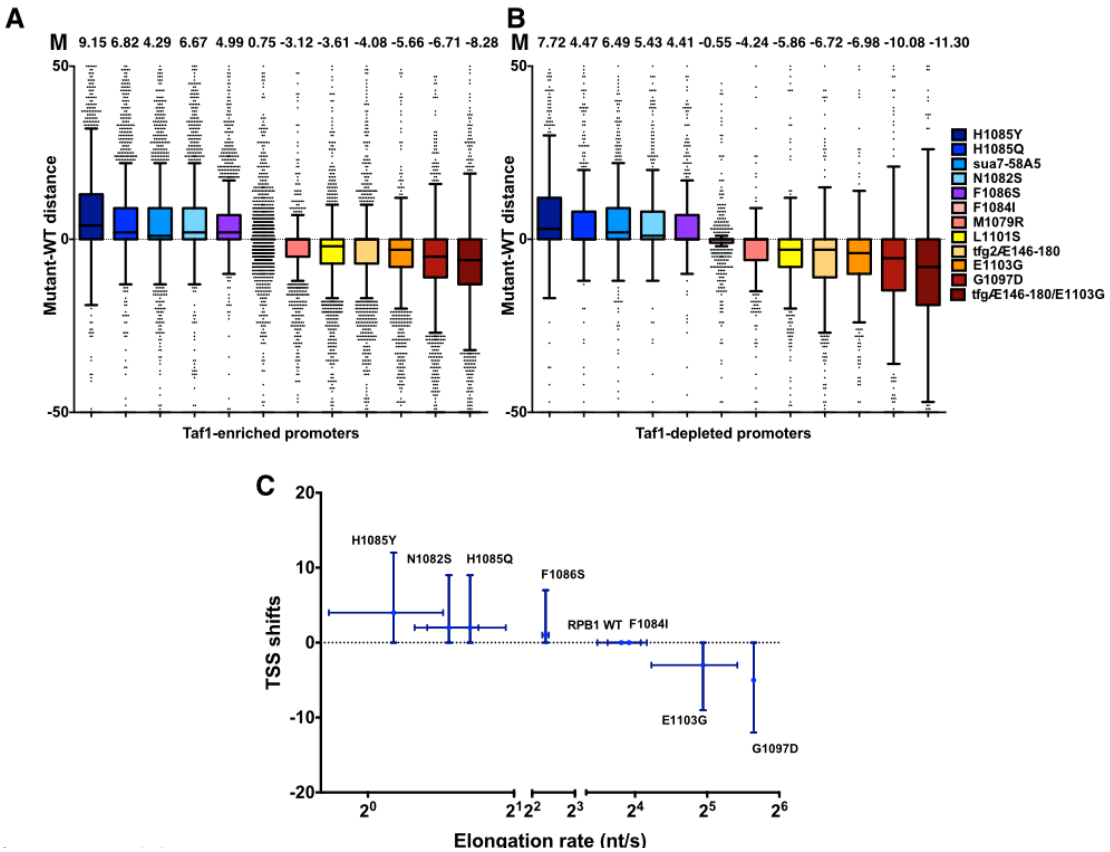


Figure 4-18. TSS shifts by Pol II activity/GTF mutants.
TSS shifts in promoters >10 rpm expression are shown for: **A.** Taf1-enriched promoters and **B.** Taf1-depleted promoters. Median numbers (M) are shown for each mutant on top of boxes. Mutants are ranked by their measured or estimated activity (slowest to fastest). **C.** Measured activity *in vitro* (estimate of mean maximal activity with 95% confidence interval on *x* axis) and their global TSS shifts (median and standard deviation on *y* axis) are shown as *x*-*y* plot.

Promoter architecture determines sensitivity of promoter output to Pol II catalytic mutants and GTF mutants

TSS region sequence is a major determinant for TSS utilization at TATA-containing model genes, as predicted by the scanning model. However, how promoters interact specifically with Pol II activity or GTFs, and whether or not promoter architecture determines sensitivity to initiation mutants is not well understood. One reason to consider this possibility is that promoters all have evolved different sequences and TSSs have evolved to be positioned at different distances from core promoter elements or nucleosomes, and these physical characteristics might be predicted to have effects on a scanning process. Previous studies suggested that when sequences functioning as efficient TSSs were moved to <50 nt downstream of another TATA element, they were not as efficient (Faitar et al., 2001). First, in hybrid promoters fusing *ADH1* or *HIS3* core promoters with TSS sequences from either promoter, *sua7* mutant effects were similar when TSS sequences were the same (regardless of fused upstream promoter sequence). Second, *ADH1* promoter mutant constructs with shorter or larger distance between TATA element and TSS all conferred sensitivity to an *sua7* mutant. However, it was not clear if changes in distances to TATA element influenced sensitivity to the *sua7* mutant because the data were not quantified (Faitar et al., 2001). Does the promoter architecture influence the TSS shift sensitivity towards Pol II activity/GTF mutants?

In heatmaps of TSS density differences for fast Pol II mutants compared to WT, with Taf1-depleted promoters rank-ordered by their median TSS distance from TATA

elements, we observed an apparent limit to the distance which TSSs could be shifted upstream from their normal locations (**Figure 4-19**). Many of these Taf1-depleted promoters will contain a consensus TATA element, and we see that upstream TSS shifts in initiation mutants seem to be limited at a distance proximal to the position of the core promoter, especially for promoters with TSSs in the 40−120 nt range from a consensus TATA box. The upstream barrier seems to be ~40 nt downstream of a consensus TATA element. Promoters that have core promoter-TSS distances of ~40−150 nt from median TSS to TATA element show different sensitivities to TSS shifting in upstream shifting initiation mutants dependent on distance between core promoter and TSSs. The further TSSs are from TATA element, the further upstream they shift (**Figure 4-19**).

We validated our visual inspection of different heatmaps by examining the median TSS shifts for promoters binned into quintiles depending on the distances between TSS and core promoter for WT. First, we examined Taf1-depleted/TATA-containing promoters shifts by Pol II activity/GTF mutants dependent on distances between median TSS and TATA element (**Figure 4-20**). Second, we examined Taf1-depleted promoters shifts dependent on median TSS and median Ssl2 ChIP-exo signals (**Figure 4-21**). Third, Taf1-enriched promoters were examined dependent on distance between median TSS and median Ssl2 ChIP-exo (**Figure 4-22**).

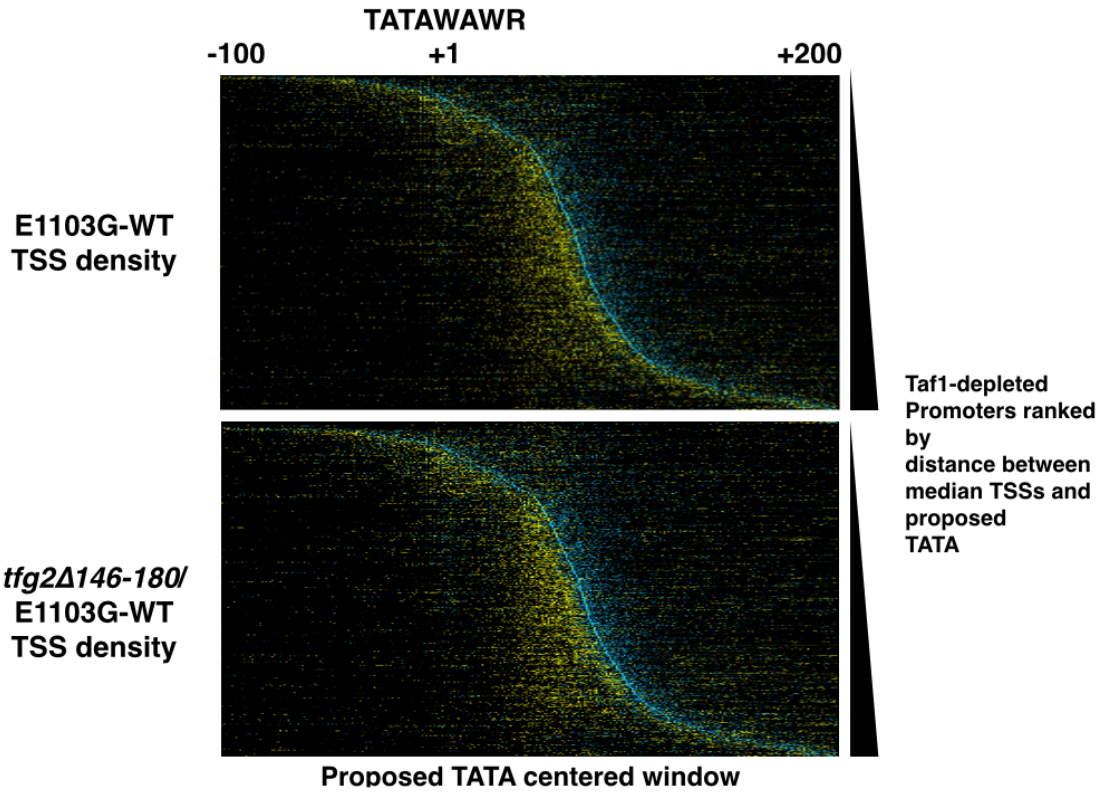


Figure 4-19. Normalized TSS density differences for fast Pol II/*tfg2* mutants in TATA element centered promoter windows.
Taf1-depleted promoters are ranked by median TSS distance to predicted TATA element. Differences on a dark background with yellow (increased usage) and blue (decreased usage) are shown for better visualization.

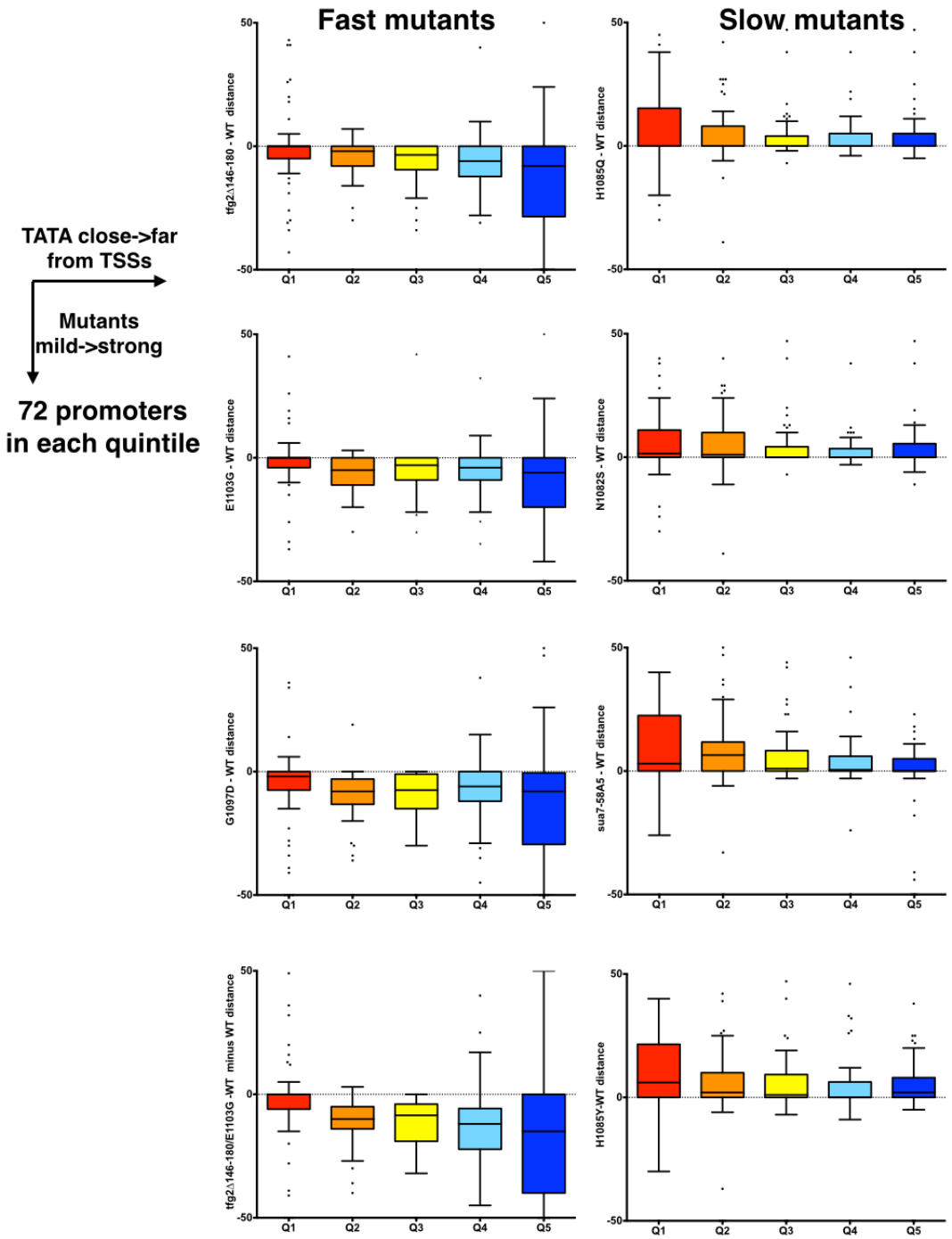


Figure 4-20. Opposite TSS shift sensitivities for two classes of initiation mutants are modulated dependent on promoter architecture (Taf1-depleted/TATA containing promoters).

Taf1-depleted/TATA containing promoters above 10 rpm expression are binned into quintiles dependent on median TSS distance to predicted TATA element. Left panel: Fast Pol II/*tfg2* mutants Right panel: Slow Pol II/*sua7* mutants.

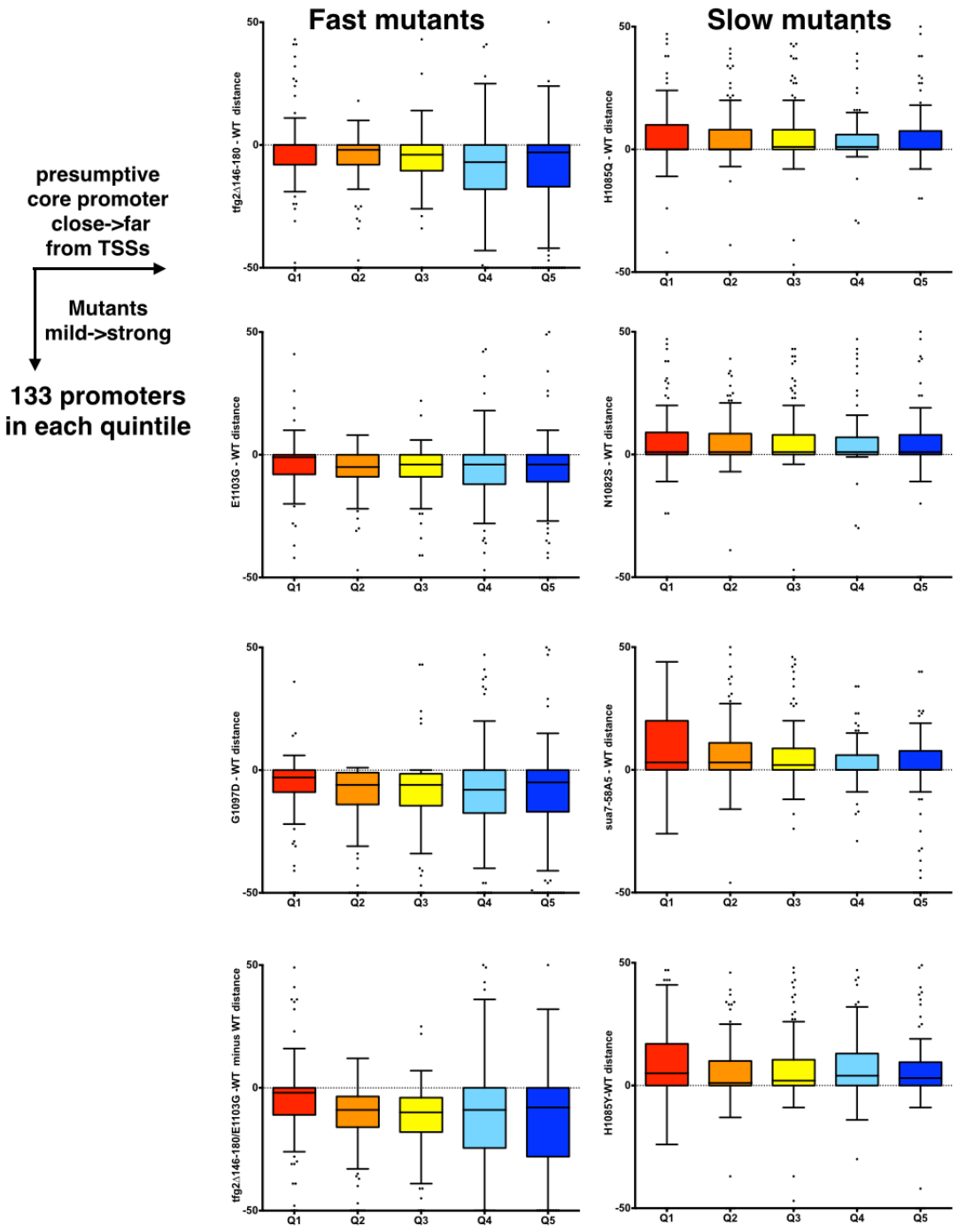


Figure 4-21. Opposite TSS shift sensitivities for two classes of initiation mutants are modulated dependent on promoter architecture (Taf1-depleted promoters). Taf1-depleted promoters above 10 rpm expression are binned into quintiles dependent on median TSS distance to median Ssl2 ChIP-exo signal. Left panel: Fast Pol II/*tfg2* mutanats, right panel: Slow Pol II/*sua7* mutants.

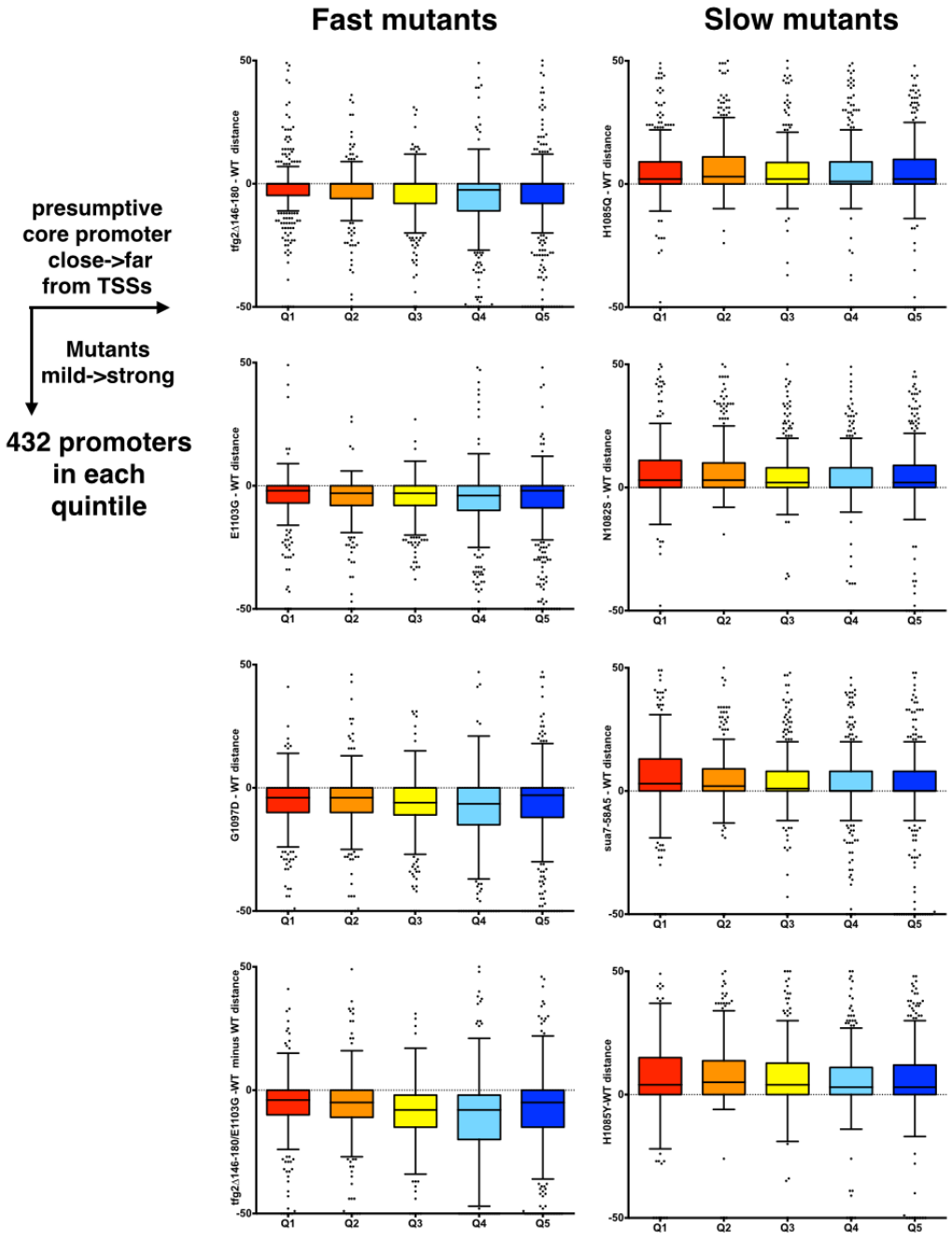


Figure 4-22. Opposite TSS shift sensitivities for two classes of initiation mutants are modulated dependent on promoter architecture (Taf1-enriched promoters). Taf1-enriched promoters above 10 rpm expression are binned into quintiles dependent on median TSS distance to median Ssl2 ChIP-exo signal. Left panel: Fast Pol II/*tfg2* mutanats, right panel: Slow Pol II/*sua7* mutants.

As seen in **Figure 4-20**, **4-21** and **4-22**, fast Pol II or *tfg2Δ146-180* mutants show greater upstream shifts when TSSs are distal to the core promoter, and show smallest upstream shifts when TSS are proximal to the core promoter, an observation most apparent in Taf1-depleted/TATA-containing promoters. In Taf1-enriched promoters, the difference pattern can be seen in Q1-Q4 promoters. Conversely, slow Pol II or *sua7-58A5* mutants show larger downstream shifts when TSSs are proximal to the core promoter. This trend was observed in Taf1-depleted/TATA-containing promoters most obviously, with mutants showing the strongest overall downstream shifting TSS effects, *sua7-58A5* and H1085Y, showing the strongest pattern based in CP-TSS distances (**Figure 4-20**). However, the trend is less apparent for Taf1-enriched promoters (**Figure 4-22**). To explain these observations, we propose there is a distance-based constraint on optimal initiation efficiency determined by the position of the core promoter (**Figure 4-23**). The mutants that shift TSS upstream are limited by the upstream edge of an "initiation efficiency window" delimited by the core promoter, and won't be able to shift TSSs as far as for other promoters with TSSs evolved to be at different distances. Conversely, while at promoters where TSSs have evolved to be far from core promoter element, an upstream shift in TSS usage will be favored position in terms of initiation efficiency, thus these promoters show a greater extent in upstream shifts. On the other hand, mutants that shift TSSs downstream favor downstream shifting in the promoters with TSSs close to the core promoter because the direction of shift moves TSSs into a more optimal position (**Figure 4-23**).

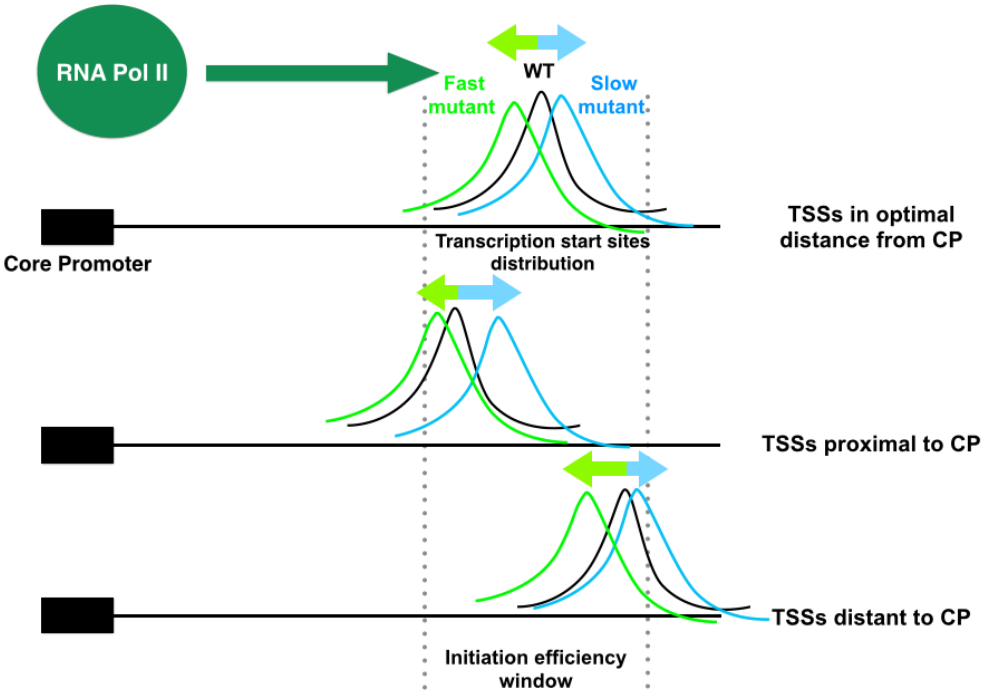


Figure 4-23. Optimal initiation efficiency model based on distance-based constraints determined by core promoter (CP).

The observation of different promoter sensitivities to TSS shifting mutants led us to suspect that initiation factors may modulate the efficiency window such that overall initiation efficiency might be altered if TSSs are shifted to more or less optimal positions within an efficiency window. While the promoter region sequence context evolves to allow a particular level of initiation, TSS positioning within an initiation window may determine the overall output of a promoter. Disruption of this relationship by altering initiation factors may lead to changes in various initiation properties including transcription output (gene expression). We therefore asked if gene expression changes in Pol II mutants showed any correlation with promoter architecture. We deployed the quantile normalization technique (Bolstad et al., 2003) on reads per million (rpm) mapped reads/promoter for WT/mutant libraries to minimize read coverage depth differences and erroneous differential effects. We calculated mutant/WT ratio of the quantile-normalized rpm for all promoters in each TSS library, grouping promoters into quintiles by WT distance between median TSS and TATA (Taf1-depleted/TATA-containing promoters) or estimated core promoter element position (median Ssl2 ChIP-exo). Surprisingly, we observed interesting trends of gene expression changes that are opposite in fast and slow mutants, and dependent on their core promoter distances to TSSs (**Figure 4-24**, **4-25**, **4-26**). These trends emerge from the background of gene expression changes (levels and noises) that are caused by many reasons.

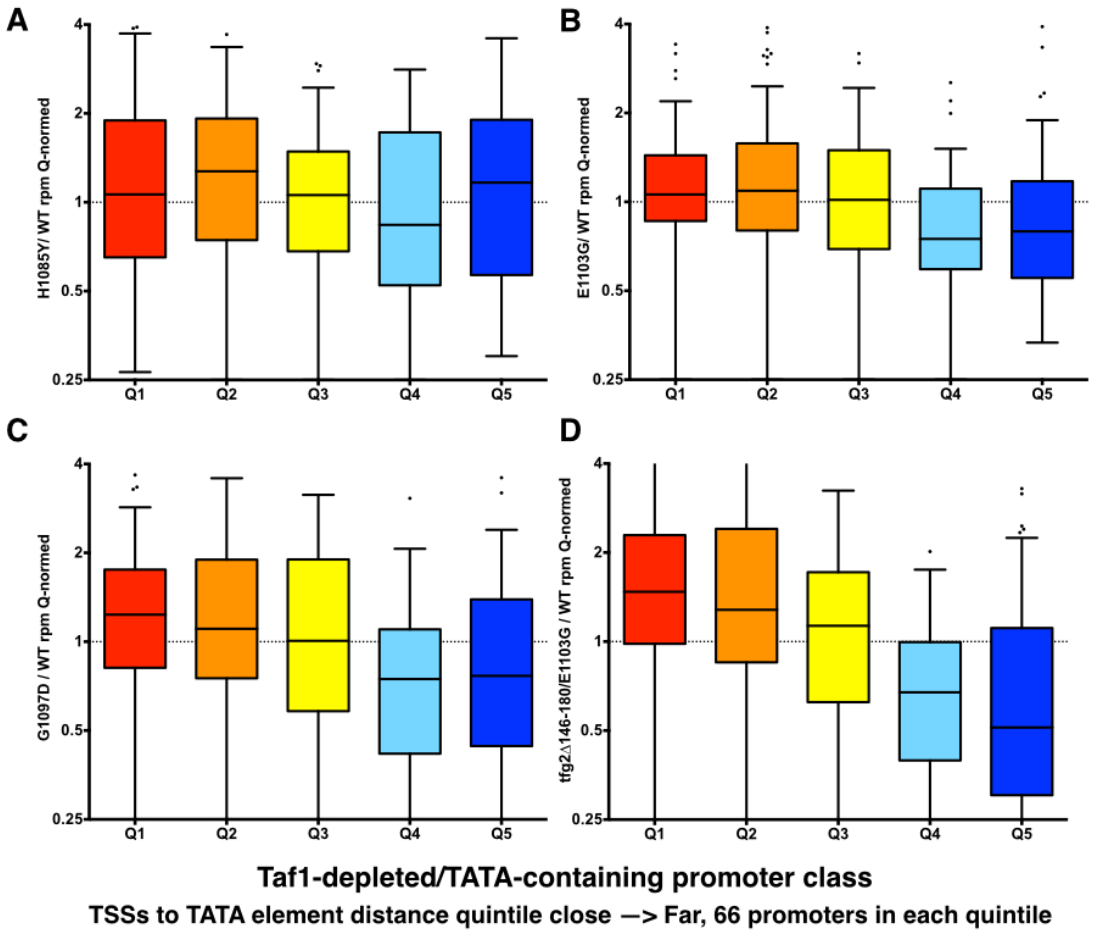


Figure 4-24. Expression changes in Pol II activity/GTF mutants show trends dependent on promoter architecture (Taf1-depleted/TATA-containing promoters). Taf1-depleted/TATA containing promoters above 10 rpm expression are binned into quintiles dependent on median TSS distance to predicted TATA element. Expression changes are calculated by ratio of mutant quantile normalized rpm/WT quantile normalized rpm for **A.** slow Pol II mutant H1085Y, fast Pol II mutants **B.** E1103G, **C.** G1097D and Pol II/*tfg2* double mutant **D.** *tfg2*Δ*146-180*/E1103G.

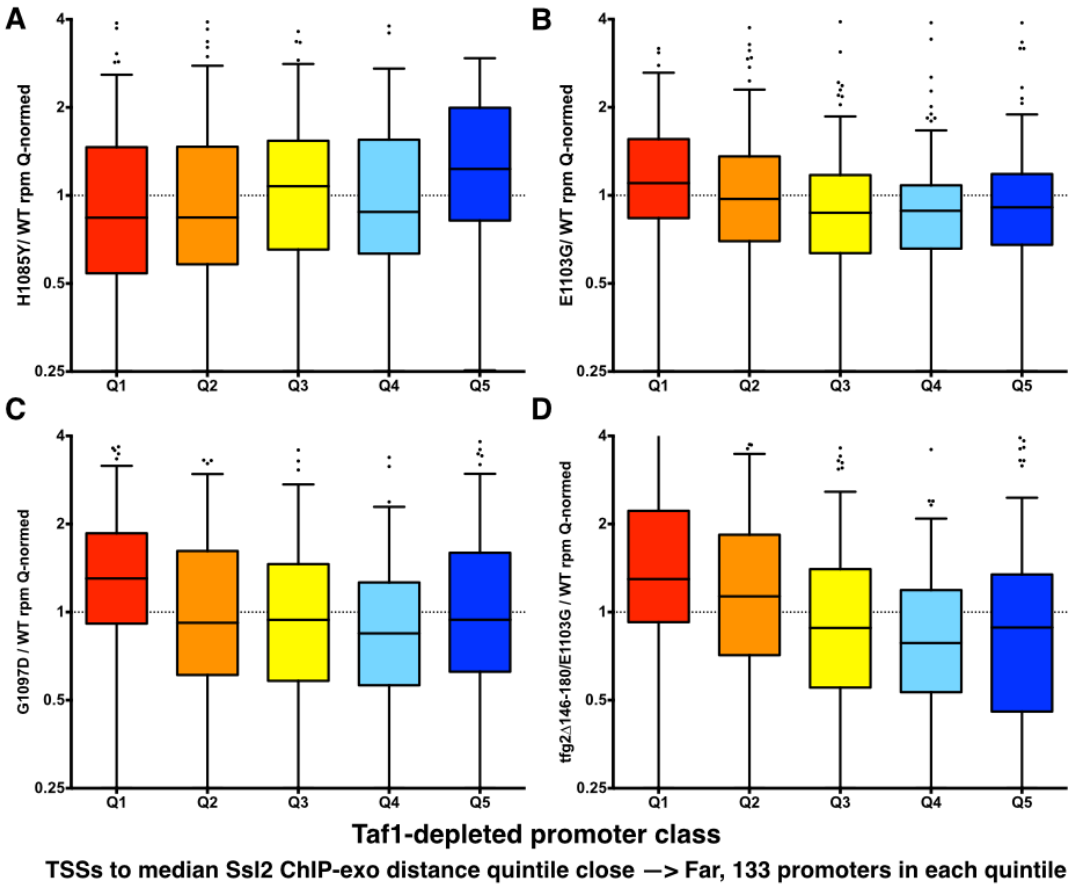


Figure 4-25. Expression changes in Pol II activity/GTF mutants show trends dependent on promoter architecture (Taf1-depleted promoters).
Same as **Figure 4-24**, but for Taf1-depleted promoters above 10 rpm expression binned into quintiles dependent on median TSS distance to median Ssl2 ChIP-exo signal.

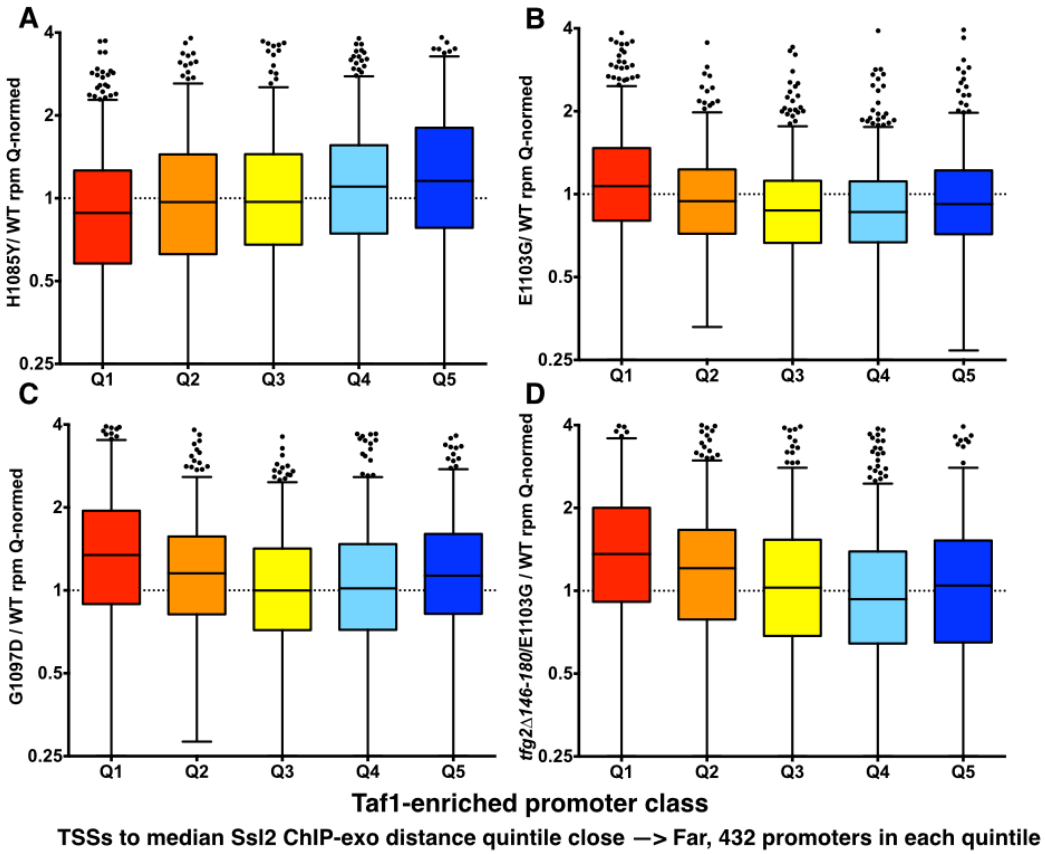


Figure 4-26. Expression changes in Pol II activity/GTF mutants show trends dependent on promoter architecture (Taf1-enriched promoters).
Same as **Figure 4-24**, but for Taf1-enriched promoters above 10 rpm expression binned into quintiles dependent on median TSS distance to median Ssl2 ChIP-exo signal.

In fast Pol II/TFIIF that shift TSS upstream in the majority of promoters, promoters that have TSSs closest to core promoters show majority of promoters upregulated. The further TSSs are from core promoter elements, the less promoters are upregulated and this trend of expression changes is present both in Taf1-enriched and Taf1-depleted promoter classes (**Figure 4-24**, **4-25**, **4-26**). Contrarily, slow Pol II/TFIIB mutants show a gene expression change trend that is the opposite of that of fast Pol II/TFIIF mutants for Taf1-enriched promoters (**Figure 4-26**). Taf1-depleted promoters do not show a clear trend for slow mutant expression changes (**Figure 4-24**, **Figure 4-25**).

We recognize that the trends of expression changes conferred by Pol II activity mutants and GTF mutants we observed might stem from various reasons. We asked if the trend we observe is due to differential enrichment of co-regulated promoters among promoter quintiles (Ihmels et al., 2002). Genes that cluster in expression profiling under various conditions are considered co-regulated. The expectation for the hypothesis that gene expression change patterns we observe are mainly due to co-regulated genes is: First, we may find co-regulated genes highly enriched for a certain quintile(s). Second, these co-regulated genes may show a specific regulation (up or down) in our Pol II activity/ GTF mutants regardless of their distance quintile. Third, co-regulated genes may show different gene expression change patterns (if there is any change) from the rest of the genes present in each quintile. For instance, if co-regulated genes tend to be upregulated in fast Pol II mutants but are also enriched in promoter quintiles that have a short distance between core promoter and TSSs (Quintile 1 and 2), this could contribute

to our observed expression trend. We searched for Gene Ontology (GO) annotation database for co-regulated genes (Ihmels et al., 2002) and annotated promoters in our study. We found Ribosomal Protein (RP) promoters to be distributed unevenly in distance quintiles of Taf1-enriched promoters. However, RP genes in our fast mutants do not show uniform behavior for their expression change (up regulation or down regulation) as expected of co-regulated genes. Most of RP promoters in Quintile 1 (Q1) are upregulated in fast mutants E1103G, G1097D and *tfg2*Δ*146-180*/E1103G, while the majority of RP promoters in Q2-Q4 are downregulated. When examining an example of the alternate class of Pol II mutant, the slow mutant H1085Y, we observe that most RPs appear to be down-regulated, while still exhibiting a trend in the expression effect across promoter quintiles arranged by core promoter-TSS distance (**Figure 4-27**). RP promoters appear extra sensitive to promoter architecture in this mutant because they show a stronger gene expression trend than for non-RP genes. We searched for the closely related class of co-regulated genes Ribosome Biogenesis (RiBi) genes (Jorgensen et al., 2004) in the promoters in our study, we found very low numbers of RiBi promoters in each quintile yet distributed evenly, and saw the same expression change pattern in RiBi and non-RiBi promoters as we observed in all promoters for all mutants tested.

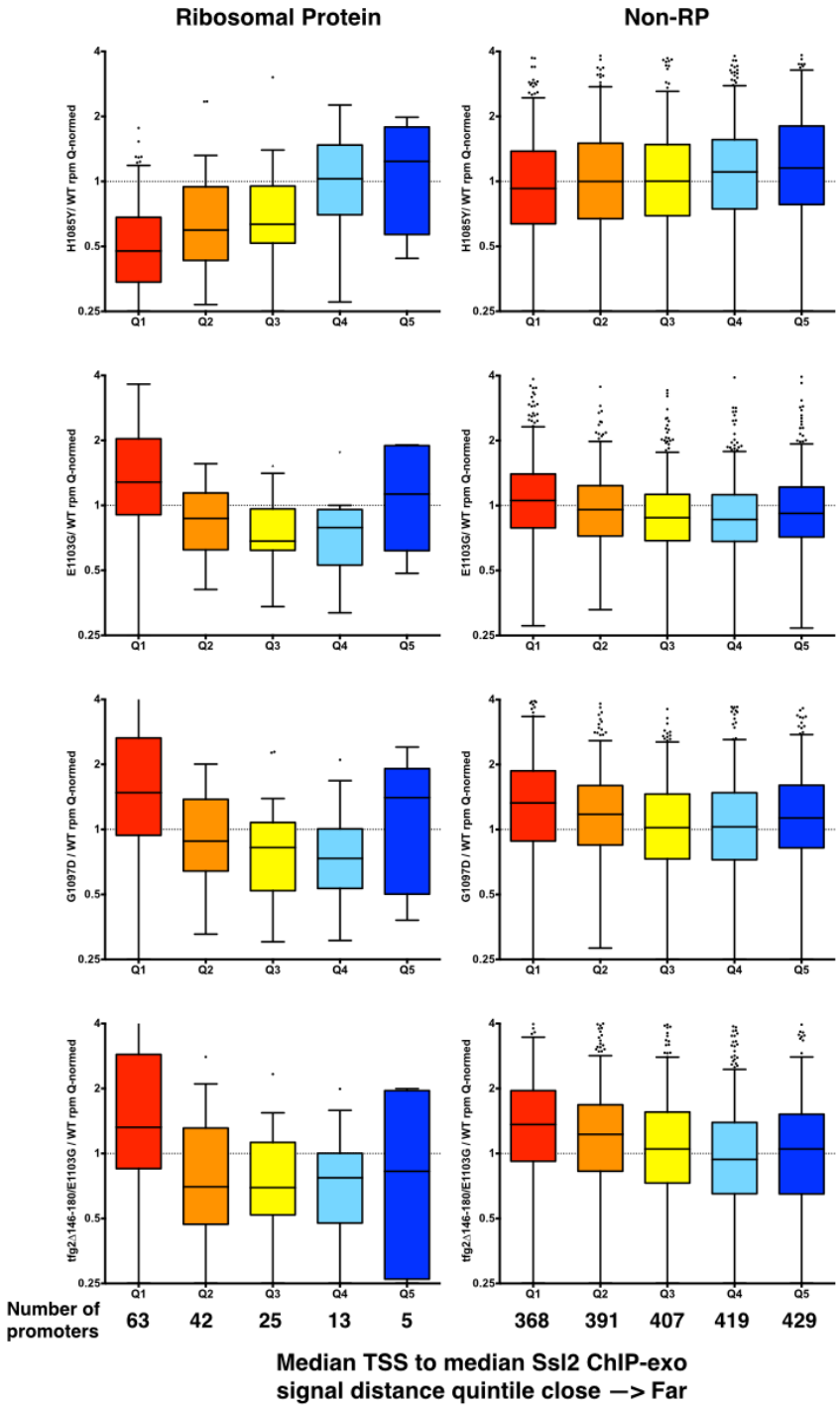


Figure 4-27. Uneven co-regulated gene enrichment is not driving force for promoter architecture modulation of sensitivity to Pol II activity.
Expression changes for Ribosomal Protein (RP) promoters (left panel) and non-RP promoters (Right panel) are shown as in **Figure 4-24** for all four mutants examined.

We recognize gene expression changes are caused by many different reasons. For example, there could be numbers of promoters that turn off gene expression in a Pol II activity/GTF mutant for physiological reasons or reasons secondary to the transcription defects in a particular mutant. However, because we can observe the trend within these various expression changes by ranking the promoters by their distances between core promoter and TSS (one key factor of promoter architecture), we suggest that evolution or promoter architecture is a contributor to gene expression. This conclusion is greatly supported by a recent publication (Lubliner et al., 2015). Pol II activity mutants are likely to alter the transcription initiation efficiency window at promoters and disrupt the relationship between Pol II scanning and initiation properties set by promoter architecture, and we propose that this has consequences for overall promoter expression.

Promoter sequence composition in Taf1-enriched and Taf1-depleted promoters

S. cerevisiae promoters are highly adenine (A) and thymine (T) rich, and different promoter classes have been shown to differ in A/T sequence enrichment in promoter regions. Previous studies have linked promoter sequence composition with nucleosomes occupancy and suggested them to be correlated with gene expression level (Hornung et al., 2012; Lubliner et al., 2013; Lubliner et al., 2015; Mavrich et al., 2008; Rosin et al., 2012; Weiner et al., 2010; Wu and Li, 2010). We examined strand specific sequence enrichment in our median TSS aligned windows by ranking promoters by expression levels and separating into promoter classes (expression deciles are based on overall ranking for all promoters) (**Figure 4-28**).

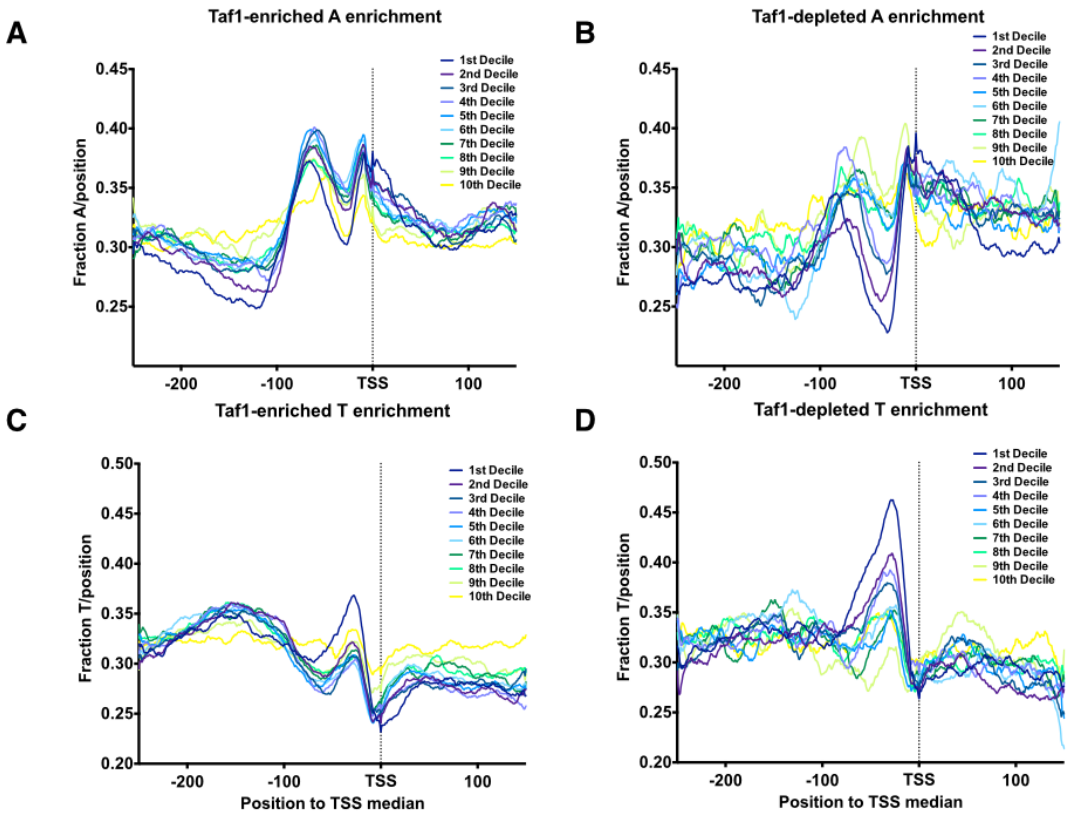


Figure 4-28. Strand-specific sequence enrichment in different promoter classes. Aggregation of A/T nucleotide normalized to number of promoter at each position in median TSS aligned promoter windows binned by expression decile (expression decile is determined by rank in all promoter; 1st: most expressed, 10th least expressed) are shown. Curves are smoothed by the LOWESS (locally weighted scatterplot smoothing) regression method. **A.** Fraction of A enrichment in Taf1-enriched promoters. **B.** Fraction of A enrichment in Taf1-depleted promoters. **C.** Fraction of T enrichment in Taf1-enriched promoters. **D.** Fraction of A enrichment in Taf1-depleted promoters.

We observed the following characteristics. First, Taf1-enriched and Taf1-depleted promoter classes have different evolved sequence skews. Second, A and T show converse strand-specific enrichment for Taf1-depleted promoters. Third, highly expressed Taf1-enriched promoters are A enriched around TSSs, show a trough just upstream, and then A enrichment upstream of the trough. Some of these have been previously observed (Lubliner et al., 2013; Wu and Li, 2010), though we note the correlation for A/T enrichment upstream of TSS much finer than previously observed, because previously promoters were separated into three expression categories (Lubliner et al., 2013). I describe strand specific sequence enrichment in Taf1-enriched and Taf1-depleted promoters as follows. Taf1-enriched promoters showed an asymmetric A enrichment in promoter region noticeable by two A peaks- the first peak was close upstream of the median TSS, while the second peak was at −80 ~ −40 nt relative to median TSS alignment. A sharp A depletion trough lay between the two A enrichment peaks, and the depth of the A depletion region weakly correlated with expression level. A wide region that was highly A depleted can be observed upstream of the second A enrichment peak. This wide region (~150 nt wide) showed a strong correlation with expression level: the higher expressed promoters showed the stronger A depletion; A depletion gradually lowered in the lower expressed promoters, with the lowest expressed promoters not A depleted over the region, instead having a similar A composition as rest of promoter window. Interestingly, T enrichment correlated with A depletion, especially at proximal upstream region of TSSs. Taf1-enriched promoters showed a first T enrichment peak immediate upstream of TSS region, with the most highly expressed

(expression decile 1) promoters showing a distinct peak. The rest of promoters in other expression deciles all show a similar enrichment level, which is lower than expression decile 1 promoters'. The −100- −200 nt (relative to median TSS) region showed a wider region of T enrichment at higher level of enrichment than the first T peak, except for first expression decile promoters. This wide T enrichment region matched location of NFR (nucleosome free region) upstream of TSSs in Taf1-enriched promoters, in consistent with (Lubliner et al., 2013; Wu and Li, 2010) (**Figure 4-28**). Conversely, Taf1-depleted promoters showed more balanced A enrichment, and the A enrichment did not show two distinct peaks as Taf1-enriched promoters did. A relatively A-depleted region close upstream of median TSSs with high correlation with expression level was observed, with the highest expressed promoters showing strongest A depletion, and lowest expressed promoters showing no A depletion, and A depletion level formed a "rainbow" dependent on expression level. Since A is a highly preferred nucleotide in TSS selection, being preferred at -8 and +1 of TSS (see below), it is possible that the A depletion region ~ −50~ −10 relative to median TSS may be one way promoters are evolved to position the TSSs in certain region. T enrichment correlated with A enrichment in Taf1-depleted promoter as well, however we observed a single T enrichment peak close upstream of median TSS in Taf1-depleted promoters, distinct from Taf1-enriched promoters. This T peak was surrounded by upstream and downstream regions relatively low T composition. T enrichment peak of Taf1-depleted promoters generally correlated with expression level, and interestingly is a mirror image of the narrow A depletion region, compensating for the A depletion, in consistent with

(Lubliner et al., 2013; Wu and Li, 2010), but showed more tight correlation in our analyses (**Figure 4-28**). T richness in this region may promote scanning because A/T base pairings thermodynamically favor melting but prevent initiation due to relative A/G depletion. A very recent publication showed that T/C richness in this region promoted transcription activity of reporter gene (Lubliner et al., 2015).

Pol II activity mutants alter TSS motif usage preferences

Local sequence is a critical factor in TSS selection and it has been shown that $Y_{-1}\underline{R}_{+1}$ motif is the dominant initiating motif in *S. cerevisiae*, as for other multisubunit RNAPs. Furthermore, in *S. cerevisiae*, highly expressed TSSs also show a strong A preference at the −8 position from the initiation position (Zhang and Dietrich, 2005). We found this consistent TSS motif in our TSS mapping as expected (**Figure 4-29**, generated by Weblogo (Crooks et al., 2004)). A is the dominant initiating nucleotide (+1) except for the lowest expressed promoters where A and G are equally preferred as starts (**Figure 4-29** C). Primarily used TSSs (*i.e.* the top used TSS) at each promoter showed strong preference for −8 A, while secondary TSSs showed a decrease in −8 A preference (**Figure 4-29** B, C). The −8 A preference was most prevalent in highly expressed promoters and was less obvious for TSSs from more lowly expressed promoters (**Figure 4-29** C, D). Narrow (TSSs from promoters with narrow spreads) and highly expressed TSSs show an additional A preference at −7~ −3 positions (**Figure 4-29** B).

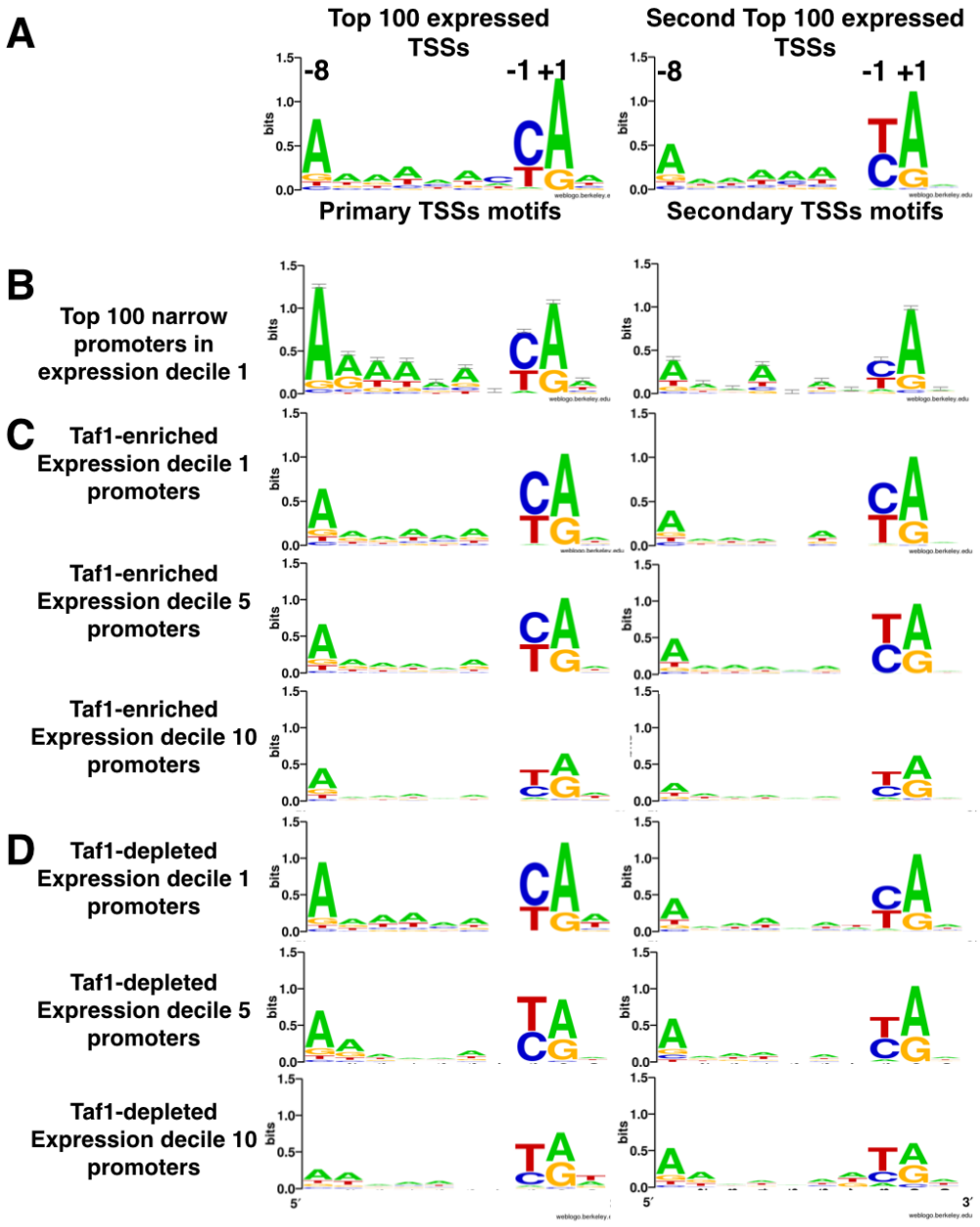


Figure 4-29. TSS motif preferences for most (primary) and second-most (secondary) used TSSs in promoters of varied expression levels.
TSS motifs for **A.** Top 100 and Top 101-200 expressed TSSs. Motifs for primary and secondary TSSs in **B.** Top 100 narrow promoters as determined by TSS spread for highly expressed promoters (expression decile 1). Motifs for primary and secondary TSSs in highly expressed (expression decile 1), medium expressed (expression decile 5), lowly expressed (expression decile 10) promoters in **C.** Taf1-enriched and **D.** Taf1-depleted promoters.

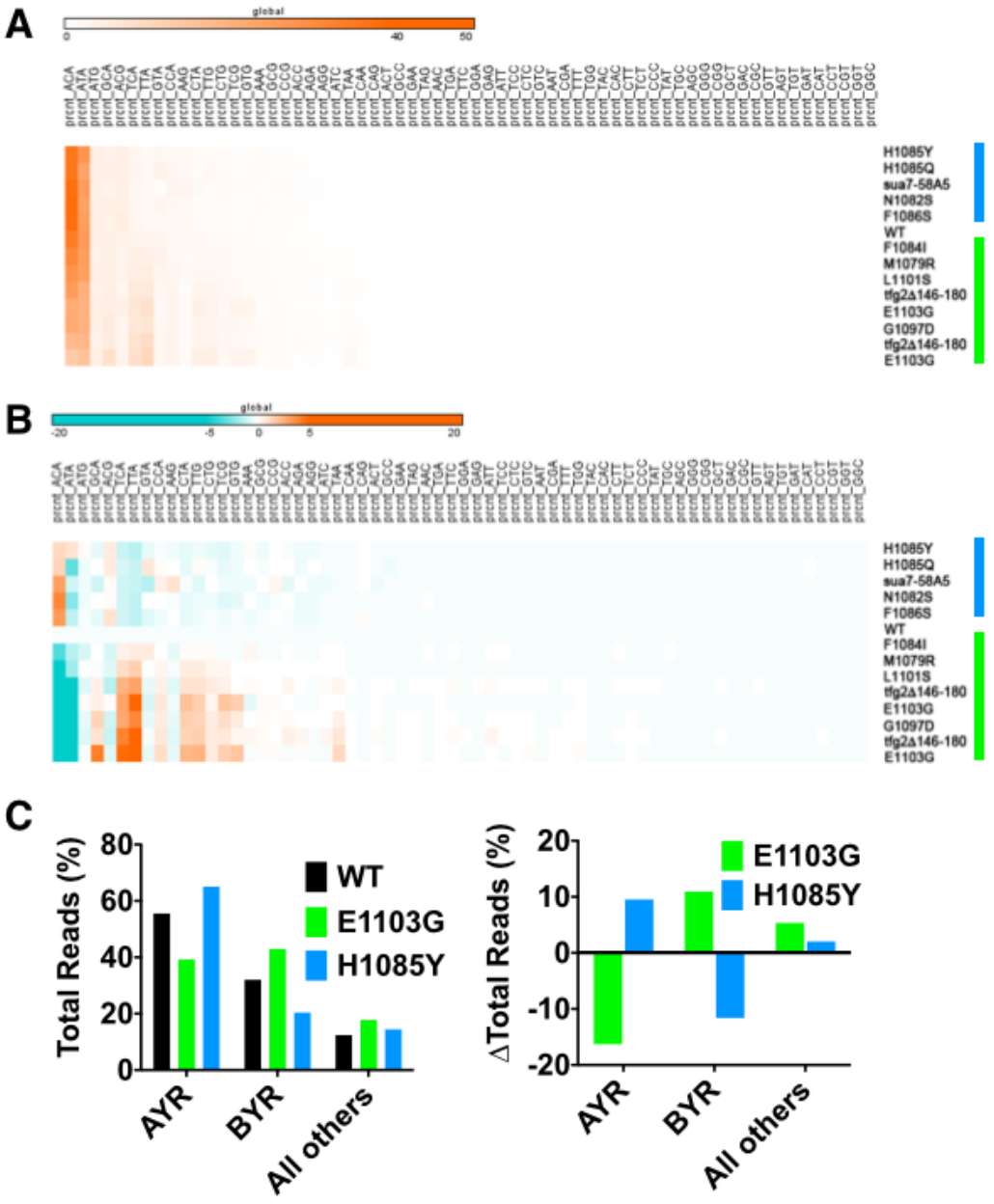


Figure 4-30. TSS motif preference changes in Pol II activity/GTF mutants.
A. TSS motif usages in Pol II activity/GTF mutants for motifs (–8, –1, +1 relative to initiating nucleotide) in the order by most used (by total TSS density) to least from left to right. Mutants are ranked by measured or estimated activity *in vitro*, slow Pol II/*sua7* mutants are marked by the blue bar, fast Pol II/*tfg2* mutants are marked by the green bar. **B.** TSS motif usage difference of data in A. **C.** $A_{-8}Y_{-1}R_{+1}$, B (not A) $_{-8}Y_{-1}R_{+1}$ and all other motif combined usage in WT, fast Pol II mutant *rpb1* E1103G and slow Pol II mutant *rpb1* H1085Y (left) and the normalized usage to WT (right).

As seen in **Figure 4-18**, Pol II activity/GTF mutants alter TSS usage distribution genome wide. The question is "how?" First we asked if these TSS shifting mutants alter TSS motif preferences. We reduced overall sequence space to examine distribution of TSS usage among the 64 motifs comprising all possible −8, −1, +1 sequences, the positions that showed distinctive sequence enrichment (**Figure 4-29**). **Figure 4-30** A shows the usages of 64 possible TSS motifs ranked by usages in WT (left: most used → right: least used). −8 A is largely preferred, three of four possible $A_{-8}Y_{-1}\underline{R}_{+1}$ motifs ACA, ATA, ATG are the top 3 most used motifs followed by a GCA motif as the 4th most used motif (**Figure 4-30** A).

Intriguingly, Pol II activity/GTF mutants alter the TSS motif usage preference (**Figure 4-30** B, C). Fast Pol II mutants and TFIIF mutants (green bar), all of which shift TSSs upstream at most promoters, show increased usages of less preferred (weak) −8 B (B=not A) TSS motifs in WT (for instance GCA, TCA, TTA), *i.e.* overall decreased specificity for −8 A. Conversely, slow Pol II mutants and TFIIB mutant (blue bar), all of which shift TSSs downstream at most promoters showed increased usages for preferred −8 A TSS motifs in WT (for instance ACA, ATA, ACG), *i.e.* increased specificity for −8 A motifs (for instance ACA, ATA, ACG) and decreased usage of less preferred −8 B TSS motifs (**Figure 4-30** B). The altered specificities for −8 A were also demonstrated for representative slow Pol II mutant H1085Y (blue) and fast Pol II mutant E1103G (green) by their usages of aggregate AYR, BYR and other (NRY) motifs in **Figure 4-30** C. Consistently with **Figure 4-30** B, slow mutant H1085Y showed increased usage of AYR motifs and decreased usage of BYR motifs, while fast mutant E1103G showed the

opposite changes in TSS motif usages (**Figure 4-30** C). We verified that TSS motif preference changes observed in **Figure 4-30** were not due to specific changes in highly expressed promoters that might dominate motif usage analyses in **Figure 4-30**. Rather, we observed TSS motif usage changes genome wide in promoter based motif usage difference analyses regardless of expression level (data not shown).

These results suggest that Pol II activity mutants change sequence specificity during TSS selection or other causes result in altered apparent sequence preferences. How might this be rationalized? One may suggest that this explains directional shifts in TSS utilization in Pol II activity mutants – because fast mutants are more catalytically active, the mutants are less "picky" about TSS motifs (lower specificity for −8 A), they are be able to utilize the less preferred or weak TSSs upstream, reducing the need to scan further. Conversely, slow mutants are catalytically defective and have greater reliance for optimal TSSs (higher specificity for −8 A, decreased usage on less preferred/weak TSS motifs). Therefore, these mutants scan further on average before initiating, leading to shift in distribution downstream (**Figure 4-31**). It is an enticing model that seems to fit the results. The above model suggests that altered sequence specificity might underlie the observations. However, an alternate model is that Pol II mutants alter efficiency across all sequence motifs, but these motifs are not uniformly present in promoters, such that altered initiation efficiency shifts initiation to different parts of promoters, where the sequence environment is different. *i.e.*, the alternative model reverses the "cause" and "observed result" of the model above.

This second model involves overall initiation efficiency at all sites. Efficiency at a certain TSS is defined by TSS usage at one position divided by sum of TSS usage at the position in question with the usage at all downstream positions for a certain promoter. During promoter scanning where Pol II is translocating directionally downstream, increased initiation activity in fast Pol II/TFIIF mutants may result in heightening the efficiency window, because probability of usage for any TSS is increased. This would result upstream shift in TSS distribution nominally independent of sequence motifs but still constrained by the nature of the sequence present. Conversely, slow Pol II and initiation defective TFIIB mutants will flatten the efficiency window due to decreased catalytic activity, leading to decreased initiation efficiency and shift in observed usage downstream (**Figure 4-32**).

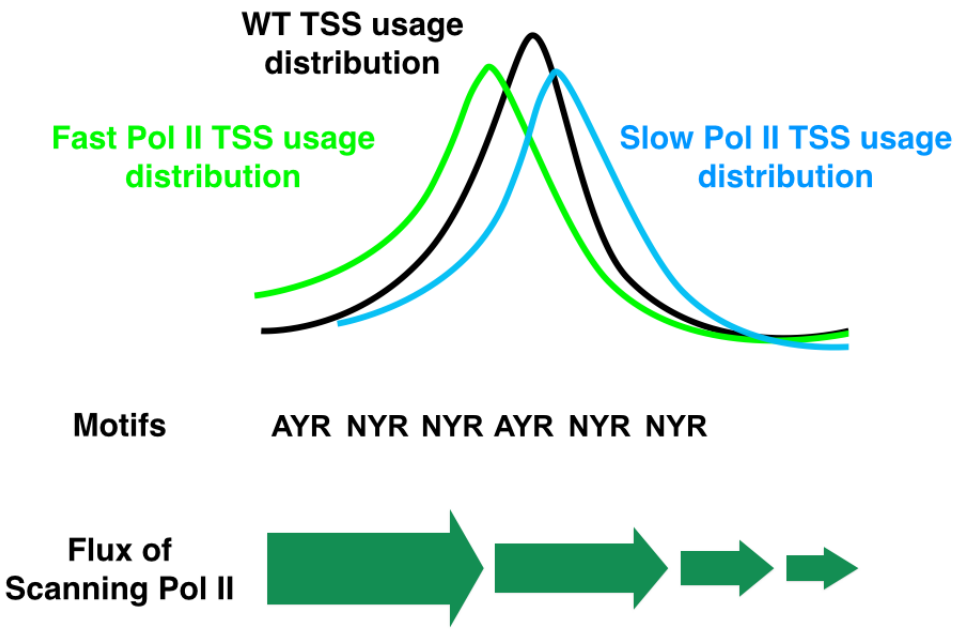


Figure 4-31. Altered TSS motif preference in Pol II activity/GTF mutants may explain polar TSS shifts.

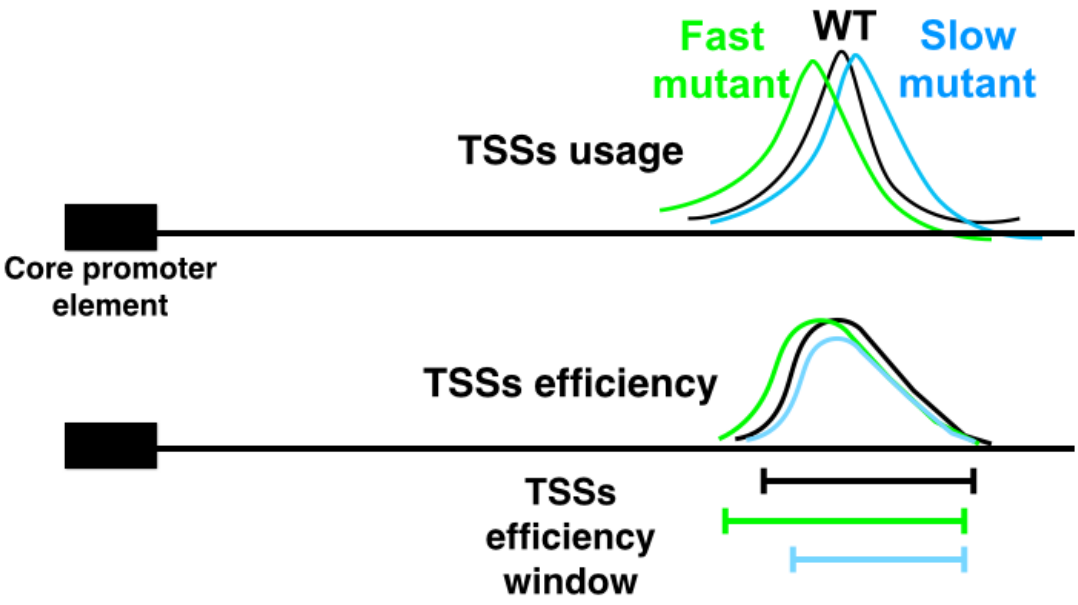


Figure 4-32. Initiation efficiency changes may lead to polar TSS shifts.

The model that proposes Pol II activity/GTF mutants alter initiation efficiency window predicts TSS usage shifting. How does it fit with altered TSS motif specificities? One possibility is that the altered efficiency windows of two classes of mutants have different TSS motif enrichments and that altered TSS motif usage is a consequence of altered TSS motif enrichment upstream and downstream of WT TSS positions. As we already know that base composition of yeast promoters is skewed around WT TSS positions (**Figure 4-28**), we extended our analysis to the availabilities of preferred −8 A TSS motifs and less preferred −8 B (not A) TSS motifs (**Figure 4-33**, **Figure 4-34**). Several observations support asymmetric localization of preferred (−8 A) and less preferred (−8 B) Y<u>R</u> motifs in TSS regions, suggesting that TSS motif usage changes in mutants can be a *consequence* of TSS shifting and not necessarily the cause. Both Taf1-enriched and Taf1-depleted promoters showed steep decrease of the most preferred TSS motif $A_{-8}C_{-1}\underline{A}_{+1}$ upstream of the median TSS position in WT that is amplified for highly expressed promoters. On the other hand, a less preferred −8 B motif with same Y<u>R</u> composition TCA (6th used motif in WT) showed a gradual and shallow decrease in availability from proximal TSS region to ~90 nt upstream region (**Figure 4-33**). Moreover, another pair of $A/B_{-8}Y_{-1}R_{+1}$ motifs showed difference in enrichment proximal to median TSSs. A preferred ATA motif (second most used motif in WT) showed a decrease upstream of median TSS, which is more prominent in highly expressed promoters and the Taf1-enriched class. Conversely, TTA motif (7th used motif in WT) enrichment showed sharp increase upstream median TSS in both promoter classes.

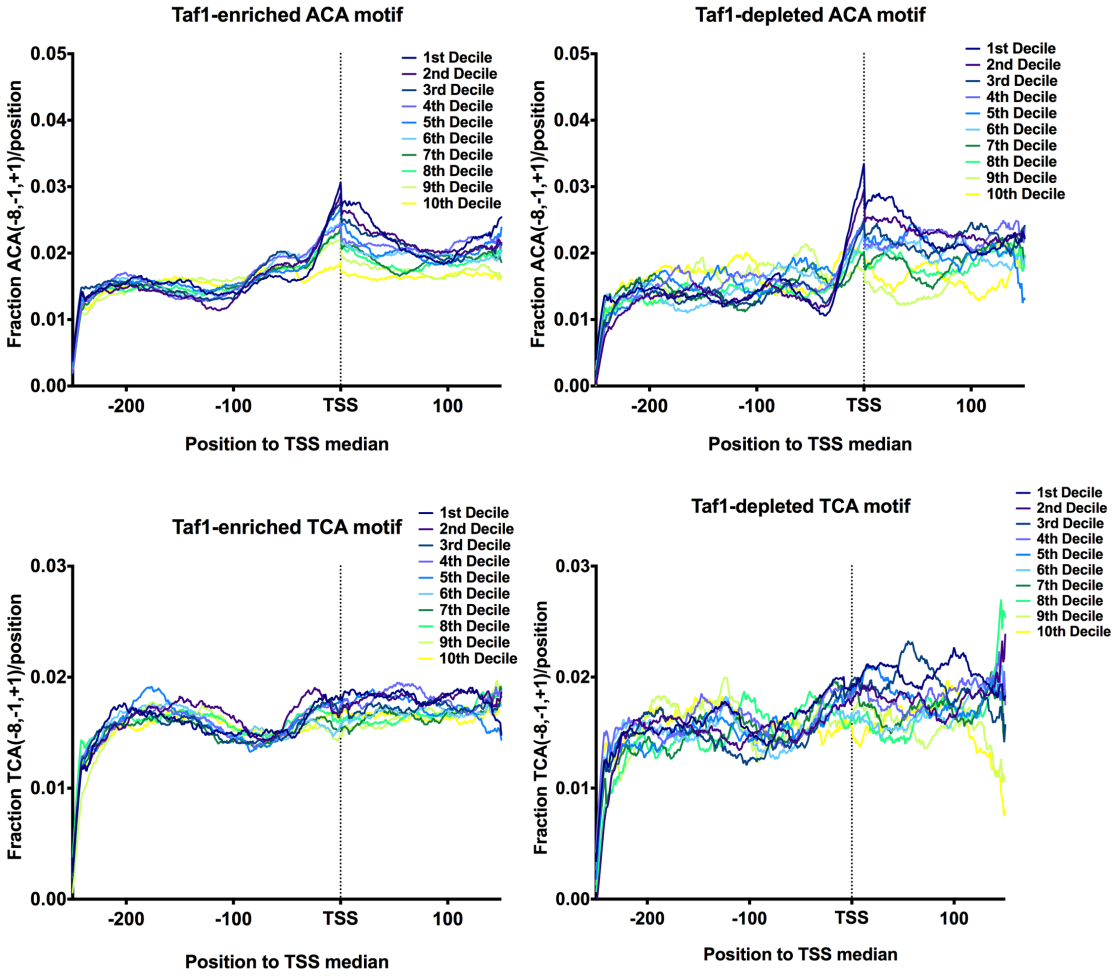


Figure 4-33. $A_{-8}C_{-1}A_{+1}$/ $T_{-8}C_{-1}A_{+1}$ sequence enrichment in median TSS aligned promoter windows.

Aggregation of $A_{-8}C_{-1}A_{+1}$/ $T_{-8}C_{-1}A_{+1}$ (a +1 position is given "1" if sequences at −8, −1, +1 position agree with the motif composition) normalized to numbers of promoter in each expression decile (expression decile is determined by ranks in all promoter; 1^{st}: most expressed, 10^{th} least expressed) is shown. Curves are smoothed by the LOWESS regression method. **A.** Fraction of $A_{-8}C_{-1}A_{+1}$ enrichment in Taf1-enriched promoters. **B.** Fraction of $A_{-8}C_{-1}A_{+1}$ enrichment in Taf1-depleted promoters. **C.** Fraction of $T_{-8}C_{-1}A_{+1}$ enrichment in Taf1-enriched promoters. **D.** Fraction of $T_{-8}C_{-1}A_{+1}$ enrichment in Taf1-depleted promoters.

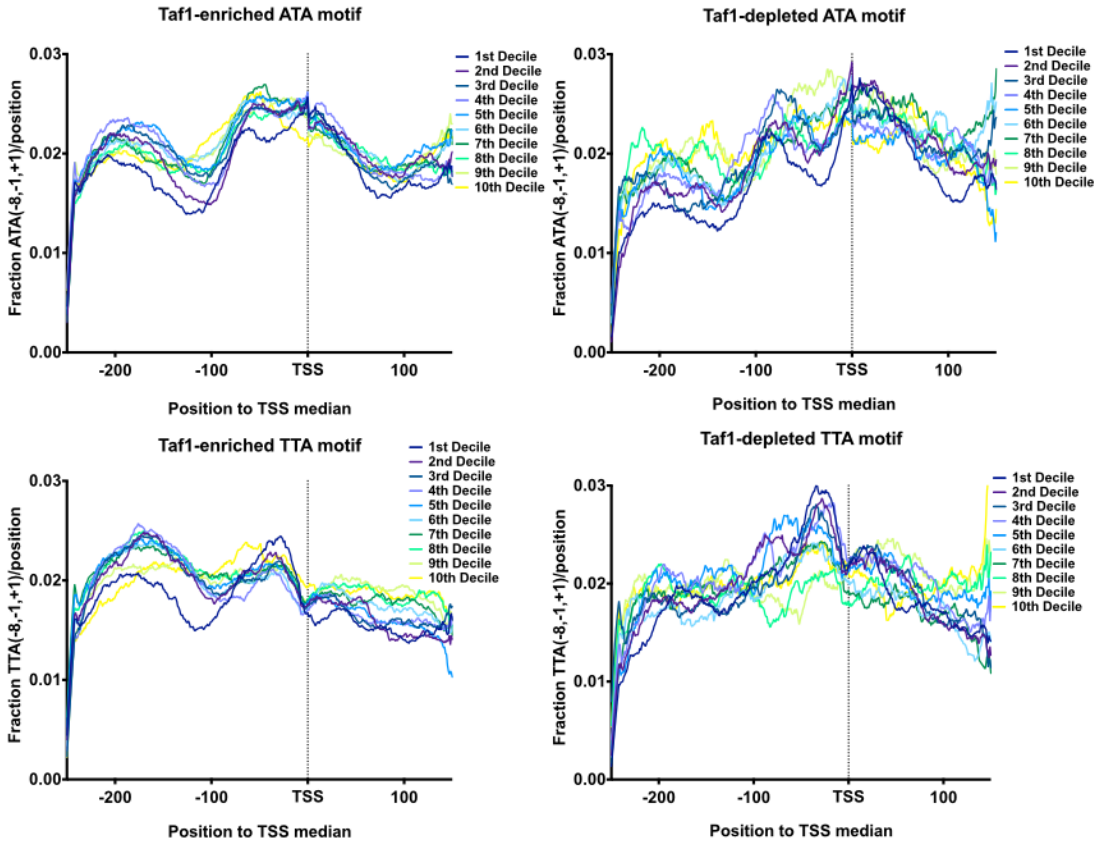


Figure 4-34. $A_{-8}T_{-1}A_{+1}$/ $T_{-8}T_{-1}A_{+1}$ sequence enrichment in median TSS aligned promoter windows.
Same as **Figure 4-33**, but for $A_{-8}T_{-1}A_{+1}$/ $T_{-8}T_{-1}A_{+1}$.

Promoter regions showing uneven enrichment of preferred −8 A motifs or disfavored TSS −8 B motifs in proximity to median TSS positions may explain the TSS motif usage changes we observed in **Figure 4-30**. Expected initiation efficiency windows in fast Pol II/TFIIF mutant (**Figure 4-32**) fall into regions where available ACA motifs (most preferred) are sharply decreased (Taf1-enriched and Taf1-depleted promoters), ATA decreased (Taf1-depleted promoters) or similar (Taf1-enriched promoters), TCA similar (Taf1-enriched and Taf1-depleted promoters) and TTA increased (Taf1-enriched and Taf1-depleted promoters). This may lead to the net difference in TSS motif usage as we observed in **Figure 4-30**. It is unclear which model or if both models are determinants for altered TSS usage distribution and different TSS motif usages observed. In order to determine the causation relationship between sequence and observed usage, we will in future implement quantitative modeling on an individual promoter basis.

Investigation of core promoter elements for TATA-less promoters

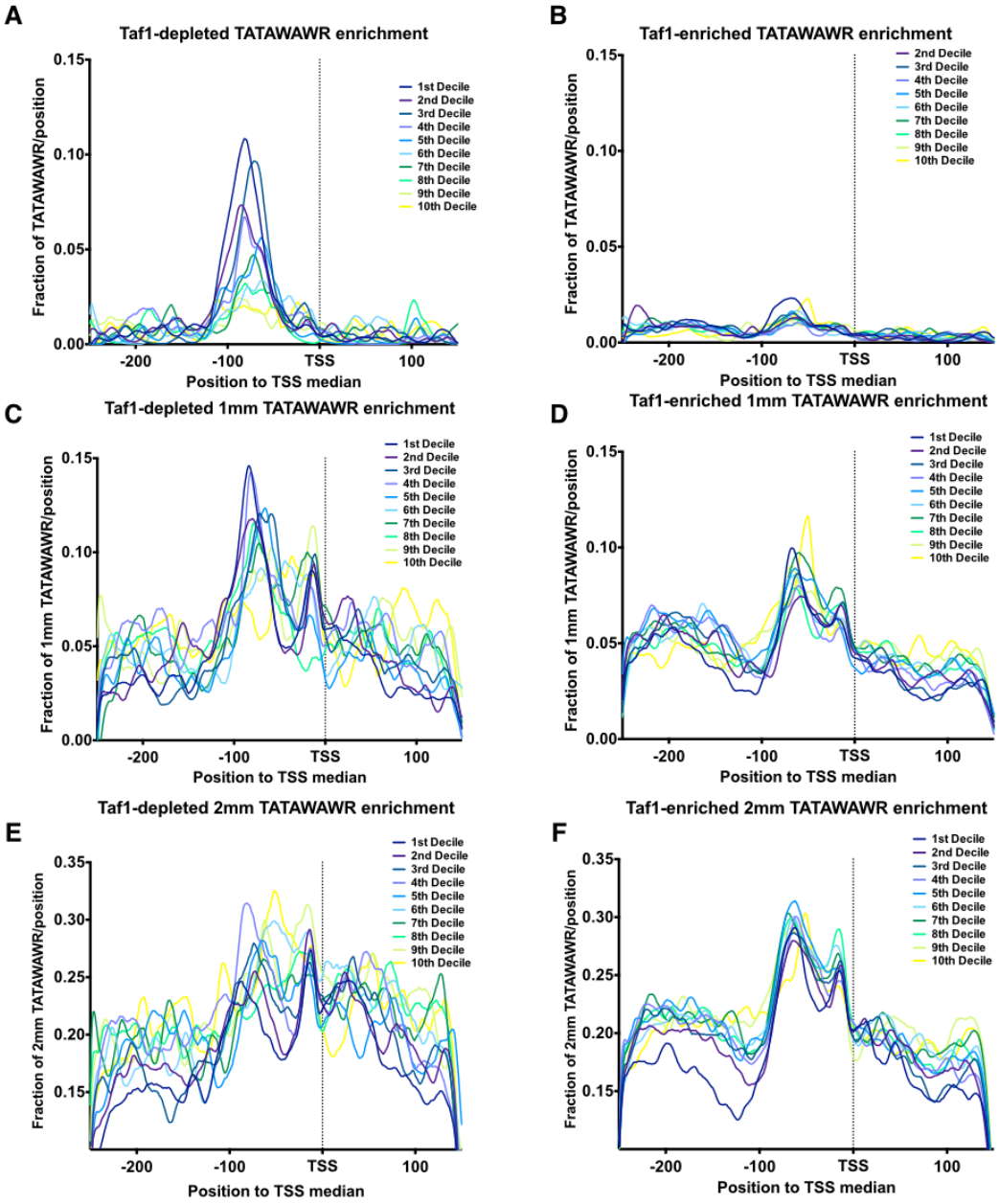
What the core promoter elements are for TATA-less promoters has been an open question. The Pugh group proposed TATA-like elements (1-3 mismatches with TATAWAWR) to be the core promoter elements as mentioned above (Rhee and Pugh, 2012), but the proposed TATA-like elements were not tested for their functionality. The Cramer group suggested the GA element could be a functional core promoter element for a set of TATA-less promoters and showed deletion or mutation at or neighbor sequences of for one GA element in a TATA-less promoter caused reduced expression (Seizl et al., 2011). More supportive evidence for GA elements were reported by others:

Wu and Li (strand specific enrichment of G capped Poly-A in NFR region of TATA-less promoters) (Wu and Li, 2010); Zhang *et al* shows that GAA repeat sequences can compensate a promoter-less reporter gene *in vivo* linked the amplification of GAA repeats to TAFs and TFIIB (Zhang et al., 2012); Zentner and Henikoff identified a long GAE motif as being enriched in regions showing altered TBP occupancy in a TBP-TATA binding modifier *mot1* mutant, (Zentner and Henikoff, 2013). Very few proposed core promoter elements for Taf1-enriched (mostly TATA-less) promoters have been tested for their functionality in driving transcription *in vivo.* The GA element in *YER175C* (a Taf1-enriched/TATA-less promoter) and variants were tested for driving expression of a reporter gene *HIS4*, and was discovered to bind to TBP *in vitro* (Seizl et al., 2011). The Hahn group showed that ribosomal proteins *RPL9a* and *RPL25* (Taf1-enriched/TATA-less promoters) promoters' expression level was not affected by purported TATA-like TBP binding site deletions (Kamenova et al., 2014)

First, we examined the enrichment of proposed promoter element by searching promoter regions aligned by median TSS for TATA-like elements or GAE (**Figure 4-35**, **Figure 4-36**). Taf1-enriched promoters did show an enrichment of 1/2 mismatch TATAWAWR (TATA-like element) at a distance from TSSs similar to where TATA boxes might be found (~−80 nt to median TSS, **Figure 4-35** D, F), however this enrichment shape was also reflective of A enrichment in this promoter class (**Figure 4-28**), not surprisingly due to the high enrichment of A in TATAWAWR and its 1/ 2 mismatch variants (**Figure 4-35** D, F). Furthermore, enrichment for TATA-like elements in the Taf1-enriched promoters were at the same level as that for Taf1-depleted

promoters, which also showed a TATA-like enrichment pattern very similar to A enrichment. A/T richness of yeast promoters and the degeneracy of a 1-2 mismatch TATAWAWR sequence indicates that they are likely to be found in all yeast promoters at possibly multiple levels. The observation of this motif at an apparent stereotypical position within ChIP-exo signal by Rhee and Pugh, similarly reflected in where the consensus TATA element is located within ChIP-exo signal (Rhee and Pugh, 2012), is intriguing but not clear how it relates to possible functionality of weak or non-consensus TATAs. Next, we searched for GA element enrichment in Taf1-enriched promoters (**Figure 4-36**). GA elements showed an enrichment peak at a distance from median TSS that is similar to the distance between TATA motif enrichment and TSSs for TATA-containing promoters (−90 ~ −40 nt), as has been observed previously using low-resolution TSS data (Wu and Li, 2010). GA elements showed semi-ideal correlation with expression similarly with enrichment of TATAWAWR in Taf1-depleted promoters (**Figure 4-35**, **Figure 4-36**). Taf1-depleted promoters did not show the enrichment peak as Taf1-enriched promoters did for GA elements (**Figure 4-36**). These results are consistent with Cramer group's hypothesis that GA element may be responsible core promoter element for a set of TATA-less promoters (Seizl et al., 2011).

Figure 4-35. TATA/TATA-like element enrichment in median TSS aligned promoter windows.
Aggregation of TATA/TATA-like elements (a +1 position is given "1" if −6 ~ +1 sequences agree with the motif composition) normalized to numbers of promoter in each expression decile (expression decile is determined by ranks in all promoter; 1st: most expressed, 10th least expressed) is shown. Curves are smoothed by LOWESS regression method. **A.** Fraction of TATAWAWR (consensus TATA) enrichment in Taf1-depleted promoters. **B.** Fraction of TATAWAWR enrichment in Taf1-enriched promoters. **C.** Fraction of 1 mismatch (mm) TATAWAWR enrichment in Taf1-depleted promoters. **D.** Fraction of 1 mismatch (mm) TATAWAWR enrichment in Taf1-enriched promoters. **E.** Fraction of 2 mismatch (mm) TATAWAWR enrichment in Taf1-depleted promoters. **F.** Fraction of 2 mismatch (mm) TATAWAWR enrichment in Taf1-enriched promoters.


A
Taf1-depleted TATAWAWR enrichment
Fraction of TATAWAWR/position
0.15
0.10
0.05
0.00
-200
-100
TSS
100
Position to TSS median
1st Decile
2nd Decile
3rd Decile
4th Decile
5th Decile
6th Decile
7th Decile
8th Decile
9th Decile
10th Decile
B
Taf1-enriched TATAWAWR enrichment
C
Taf1-depleted 1mm TATAWAWR enrichment
Fraction of 1mm TATAWAWR/position
D
Taf1-enriched 1mm TATAWAWR enrichment
E
Taf1-depleted 2mm TATAWAWR enrichment
Fraction of 2mm TATAWAWR/position
0.35
0.30
0.25
0.20
0.15
F
Taf1-enriched 2mm TATAWAWR enrichment

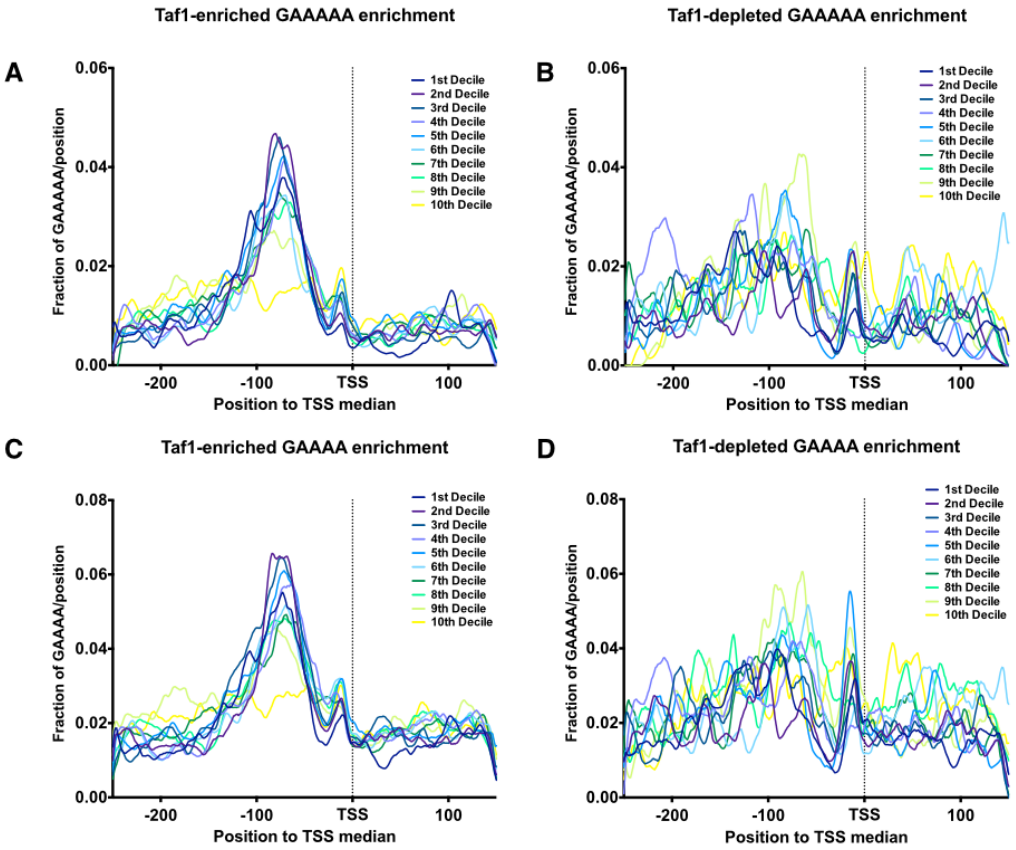


Figure 4-36. GA element enrichment in median TSS aligned promoter windows. Aggregation of GA elements (a +1 position is given "1" if −4/3 - +1 sequences agree with the motif composition) normalized to numbers of promoter in each expression decile (expression decile is determined by ranks in all promoter; 1st: most expressed, 10th least expressed) is shown. Curves are smoothed by LOWESS regression method. **A.** Fraction of GAAAAA (6 mer GAE) enrichment in Taf1-enriched promoters. **B.** Fraction of GAAAAA (6 mer GAE) enrichment in Taf1-depleted promoters. **C.** Fraction of GAAAA (5 mer GAE) enrichment in Taf1-enriched promoters. **D.** Fraction of GAAAA (5 mer GAE) enrichment in Taf1-depleted promoters.

We selected candidate promoters that were relatively highly expressed and generated plasmid constructs with their promoters regions (~500 nt fragment upstream of ATG) driving *HIS3* expression. These candidate promoters include Taf1-depleted/TATA-containing promoters with one TATA motif at stereotypical distance from TSSs, which are very likely driven by the TATA motif, and Taf1-enriched promoters with one or multiple TATA-like elements at presumptive stereotypical distance(s) or marginal distance(s) from TSS and/or a GA element (regular or extended) at presumptive stereotypical distance from TSSs (**Figure 4-37**). We constructed core promoter element mutant constructs with deletion or mutation at TATA-like (predicted by Pugh) or GA elements (found in our analyses) by site directed mutagenesis. Plasmid constructs were transformed into a yeast strain with a deleted chromosomal *HIS3*. We measured the expression of *HIS3* by northern blot experiments and compared mutant effect on expression by normalizing their expression to corresponding WT (*SED1* as loading control) (**Figure 4-38**).

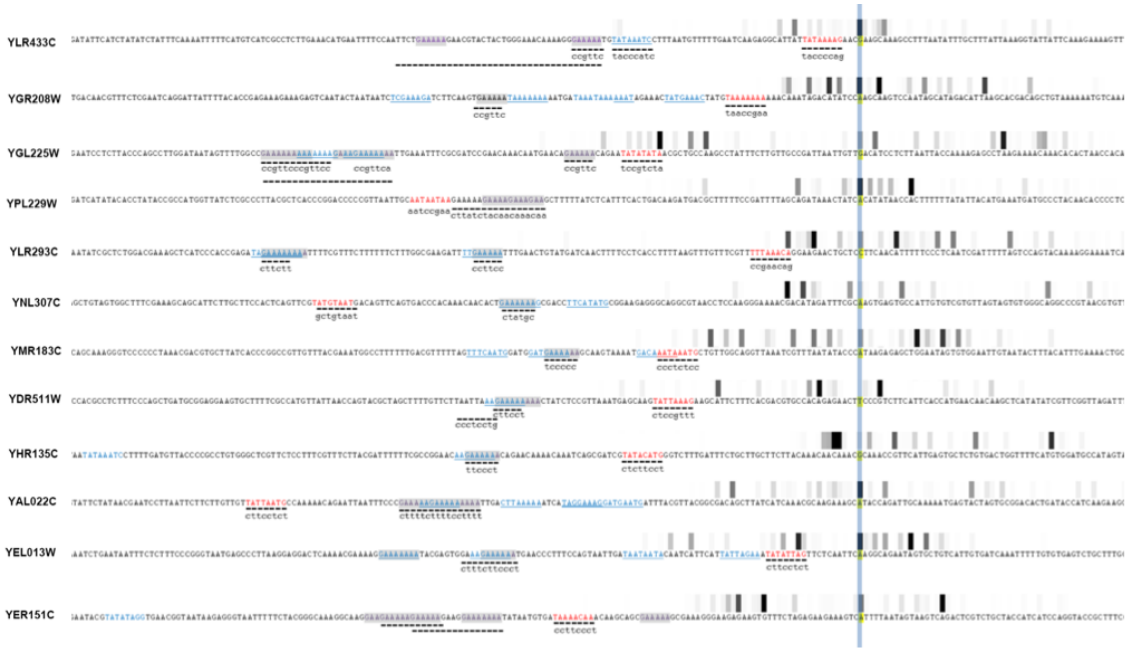


Figure 4-37. Representative Taf1-enriched promoter candidates in CPE studies. Sequences around median TSSs for each promoter candidate are aligned by median TSS in each promoter (labeled by blue bar). Predicted TATA-like elements (Rhee and Pugh, 2012) are labeled in red; other TATA-like elements are labeled blue; GA elements are highlighted in grey. Mutations at core promoter elements are shown in lower case below; deletions are labeled as dotted underlined.

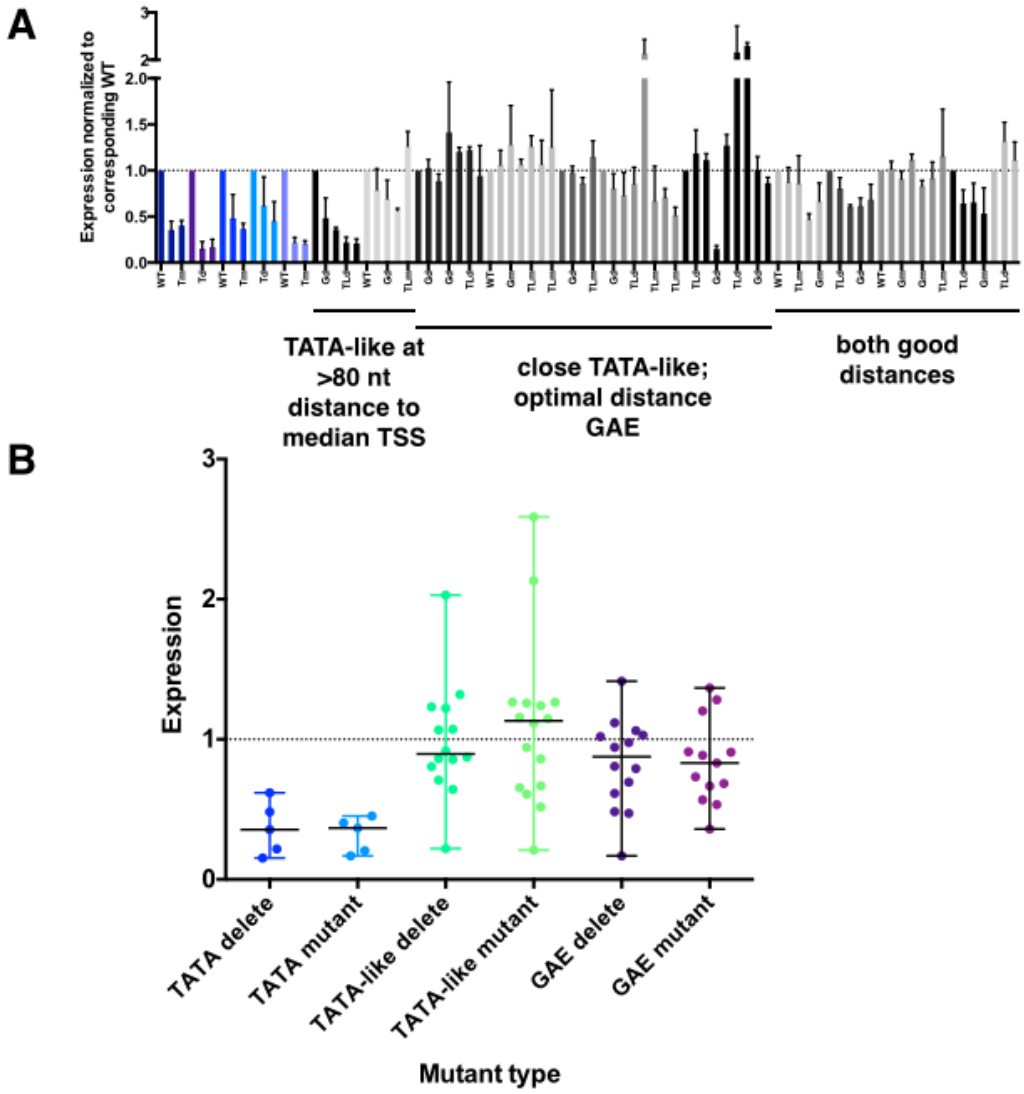


Figure 4-38. Assessment of function for TATA elements or predicted core promoter elements mutation/deletions for candidate promoters.
A. Expressions of reporter gene *HIS3* in core promoter mutant constructs are normalized to corresponding WT measured by northern blot experiments. Mean of at least three replicate experiments and shown with standard deviations as error bars. **B.** B contains same information but presented in a slightly different manner. It demonstrates the expression variances TATA-like or GA element in all promoters as scatter graph to show the magnitude of varied expression each type of promoter mutations cause.

Figure 4-38 A is a bar graph that shows the difference of expression level conferred by the mutagenesis at TATA-like elements or GA elements. Each promoter mutant expression is normalized to expression of corresponding WT in the same experiment set up (and in same blot). Deletions or mutations of GA elements in six of 11 tested Taf1-enriched/TATA-less promoters with GA element and one deletion of 12 TATA-like element containing promoters showed significantly reduced expression of the reporter gene though the reduction overall was much less than that seen for TATA deletions (**Figure 4-38**). First, three GAE deletion and two GAE mutations out of 11 promoters with GAE led to decreased expression in the magnitude of decrease similar to that caused by deletion/mutations at TATA element in tested TATA-containing promoters in the same experimental setup. This indicated that GA element might contribute to expression at two or three promoters in the tested candidates. One TATA-like deletion and one TATA-like mutation led to decreased expression in the magnitude of decrease equal to those caused by deletion/mutation at TATA element promoters (**Figure 4-38**). This experiment suggested that a set of Taf1-enriched promoters may have contributions from TATA-like elements or GA elements, but are fundamentally distinct in their contributions from consensus TATA elements. One promoter showed decreased expression in both TATA-like element and GA element mutants in the similar extent, implying that both elements might contribute to expression. Seven promoter candidates were unaffected by mutations of TATA-like or GA elements. Overall, promoter mutations in TATA-like or GA elements did not confer the magnitude of

defect in reporter gene expression seen for promoter mutations in consensus TATA elements.

Genome-wide nucleosome mapping

A model in the field proposes that nucleosome acts as a barrier for Pol II accessing template DNA thus an inhibitory role in transcription (Bondarenko et al., 2006; Rhee and Pugh, 2012; Saunders et al., 2006). It is suggested that Pol II must overcome nucleosome barriers during transcription by shifting or evicting nucleosome partially or completely to allow transcription with mechanisms proposed previously (Bednar et al., 1999; Jin et al., 2010; Kulaeva et al., 2009; Kulaeva et al., 2010). It is not hard to imagine promoter sequence context may widely regulate all initiation properties, including Pol II scanning (promoter distance, initiation efficiency), TSS selection (TSS motifs, TSS shape), initiation factor binding, promoter region nucleosome positioning etc. Nucleosomes have specific relationships to yeast promoter classes (discussed in Intro and in **Figure 4-28**). First, the Taf1-enriched class of yeast promoters has stereotypically positioned +1 nucleosomes as well as a NDR/NFR over a region of the core promoter. The +1 nucleosome overlaps with the TSSs in some promoters, raising the question of how initiation happens at these promoters. What is the relationship of TSS selection and nucleosome positioning? Does one affect the positioning of the other (**Figure 4-39)**? To further understand relationships between TSS selection and nucleosome positioning, we mapped nucleosomes in Pol II activity mutants H1085Y (slow) and E1103G (fast) genome wide. We reasoned that if Pol II mutants shift TSSs,

nucleosomes might also shift, if nucleosomes need to transiently slide to allow initiation at TSSs within nucleosomes.

We performed nucleosome position mapping in WT and one representative fast mutant, *rpb1* E1103G, and one representative slow mutant, *rpb1* H1085Y, using paired-end Illumina Truseq sequencing of Micrococcal Nuclease (MNase) digestion protected DNA fragments. See **Methods** for experimental detail. See **Table 4-5** for sequencing reads numbers in nucleosome sequencing. We obtained MNase resistant DNA fragments corresponding to mono-nucleosomes at multiple MNase digestion levels to mitigate bias in the experiment due to differential digestion levels between samples. Because we performed paired-end sequencing, we were able to map both ends of nucleosome-protected DNA fragments, **Figure 4-40** present the fragment sizes distribution in each library. Majority of fragments were within the 140-170 nt length range as expected.

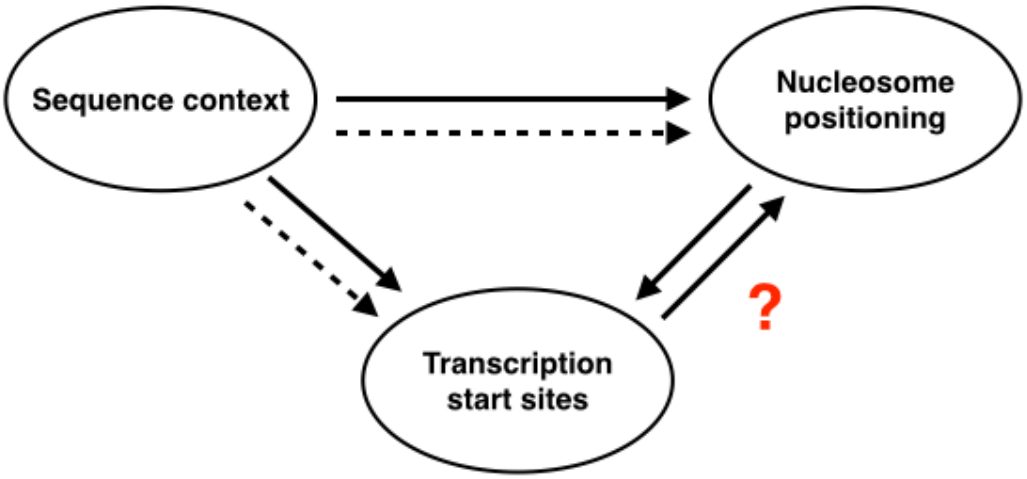


Figure 4-39. What is the relationship of TSS selection and nucleosome positioning?

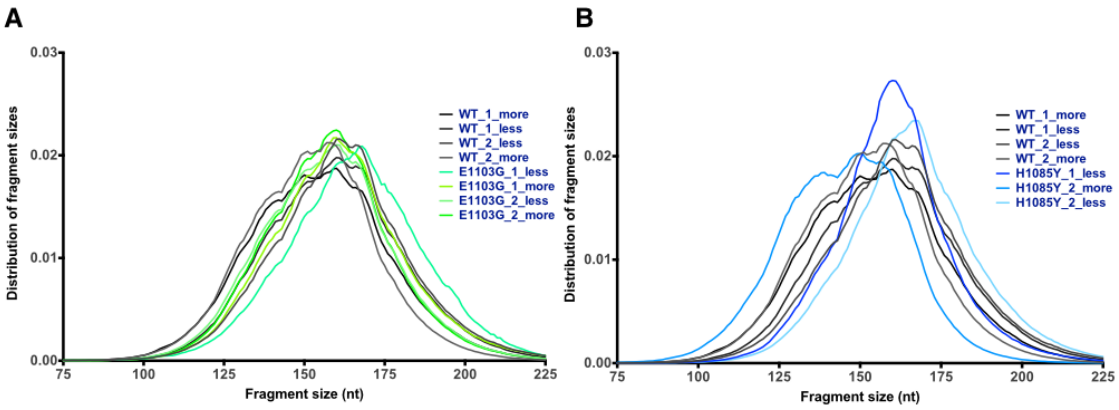


Figure 4-40. Fragment size distributions in WT/Pol II activity mutant nucleosome sequencing libraries.
Fragment sizes are calculated by difference of two ends of each fragment detected by paired-end sequencing. Fraction of each fragment sizes are plotted for **A.** WT and fast mutant E1103G libraries and **B.** WT and slow mutant H1085Y libraries.

Table 4-5 Nucleosome mapping reads

Library	Strain	Digestion	Uniquely mapped reads	Midpoint numbers
r3s1	WT	less	42,739,017	31,488,512
r3s2	WT	more	57,640,449	42,448,757
r3s5	WT	less	53,001,306	38,198,414
r3s6	WT	more	37,777,935	28,555,739
r3s9	E1103G	less	51,346,028	35,229,796
r3s10	E1103G	more	40,156,298	31,166,806
r3s11	E1103G	less	38,388,689	28,074,037
r3s12	E1103G	more	49,237,724	36,809,347
r3s13	H1085Y	less	40,045,048	30,064,946
r3s14	H1085Y	more	32,955,563	23,612,024
r3s15	H1085Y	more	46,055,013	36,354,568
r3s16	H1085Y	less	36,894,162	28,348,199

Nucleosome positions in our data were consistent with those previously observed (Field et al., 2009; Hughes et al., 2012; Kaplan et al., 2009; Lee et al., 2007; Rhee and Pugh, 2012; Rosin et al., 2012; van Bakel et al., 2013; Weiner et al., 2010; Wu and Li, 2010; Yuan et al., 2005). A number of yeast promoters have TSSs upstream of well-positioned nucleosomes, and this organization is much more apparent for Taf1-enriched promoters. As presented in **Figure 4-41**, Taf1-enriched promoters have stereotypically positioned +1 nucleosome and the midpoints are mapped to downstream of TSS. However, the midpoints very close to TSS (<~70 nt distance) indicate that some +1 nucleosomes greatly overlap with TSS positions, leading to predictions that these nucleosomes are shifted or evicted partially or completely to allow transcription (Bednar et al., 1999; Jin et al., 2010; Kulaeva et al., 2009; Kulaeva et al., 2010). A ~200 nt wide region upstream of TSSs that is usually referred to as NDR/NFR (nucleosome depleted/free region) is observed, but with a "0" (between −1 and +1 nucleosome) fragile nucleosome with lower occupancies only apparent at lower MNase digestion levels (Brogaard et al., 2012; Henikoff et al., 2011; Xi et al., 2011). Taf1-depleted promoters lack stereotypical nucleosome positioning and NFRs as has been observed previously (Kaplan et al., 2009; Tirosh and Barkai, 2008; Wu and Li, 2010) (**Figure 4-41**), consistent with the model that initiation factors and +1 nucleosomes may compete at these promoters (Rhee and Pugh, 2012). This model and others suggested that nucleosome occupancy is an important determinant for gene expression by playing an inhibitory role in gene expression (Field et al., 2009; Hornung et al., 2012; Rhee and Pugh, 2012; Rosin et al., 2012), predicting higher expressed promoters to show lower

occupancy of nucleosome especially in Taf1-depleted promoters (Rhee and Pugh, 2012). We examined W (A/T) enrichment that is shown to be disfavored sequences for nucleosome positioning (Kaplan et al., 2009; Nelson et al., 1987) in promoters binned into deciles dependent on expression level (expression decile is determined in all 6044 promoters) (**Figure 4-42**). We did not observe a direct correlation predicted by the model, W enrichment does not show any distinct enrichment or depletion correlated by expression level; similarly nucleosome occupancies do not show a strong correlation with expression level. Instead, nucleosome occupancies show a positive weak correlation with expression level in Taf1-enriched promoters (**Figure 4-42**). An explanation for this result is that highly expressed promoters have more stereotypical chromatin structure and tighter regulation of nucleosome positioning. Rather than a direct correlation with expression level, nucleosome positioning may correlate with another characteristic of promoter architecture, which is a main determinant of expression.

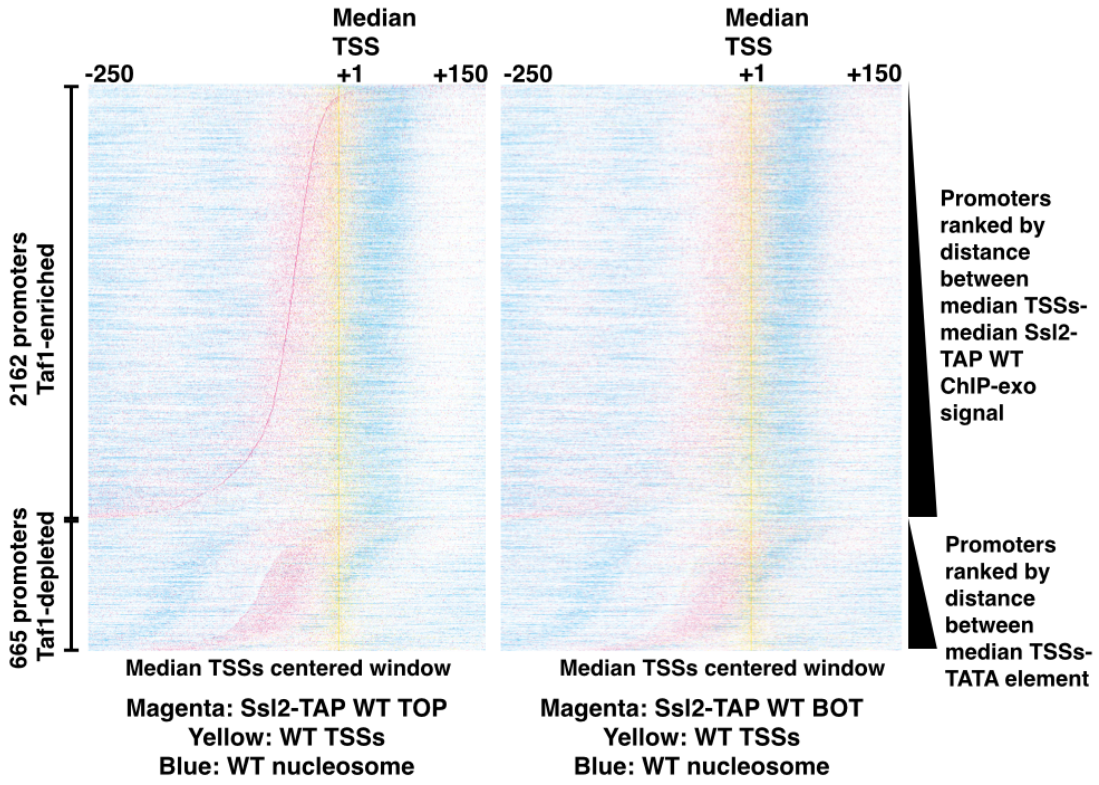


Figure 4-41. Nucleosome positioning in median TSS aligned promoter windows. Promoters are sorted by Taf1-enriched and Taf1-depleted and ranked by: distance between median TSS to median Ssl2 ChIP-exo signal for Taf1-enriched promoters or distance between median TSS to TATA element for Taf1-depleted promoters. Normalized densities of TSS and ChIP-exo (left: TOP strand; right: BOTTOM strand) are shown in yellow and magenta correspondingly. Nucleosome midpoints are shown in blue.

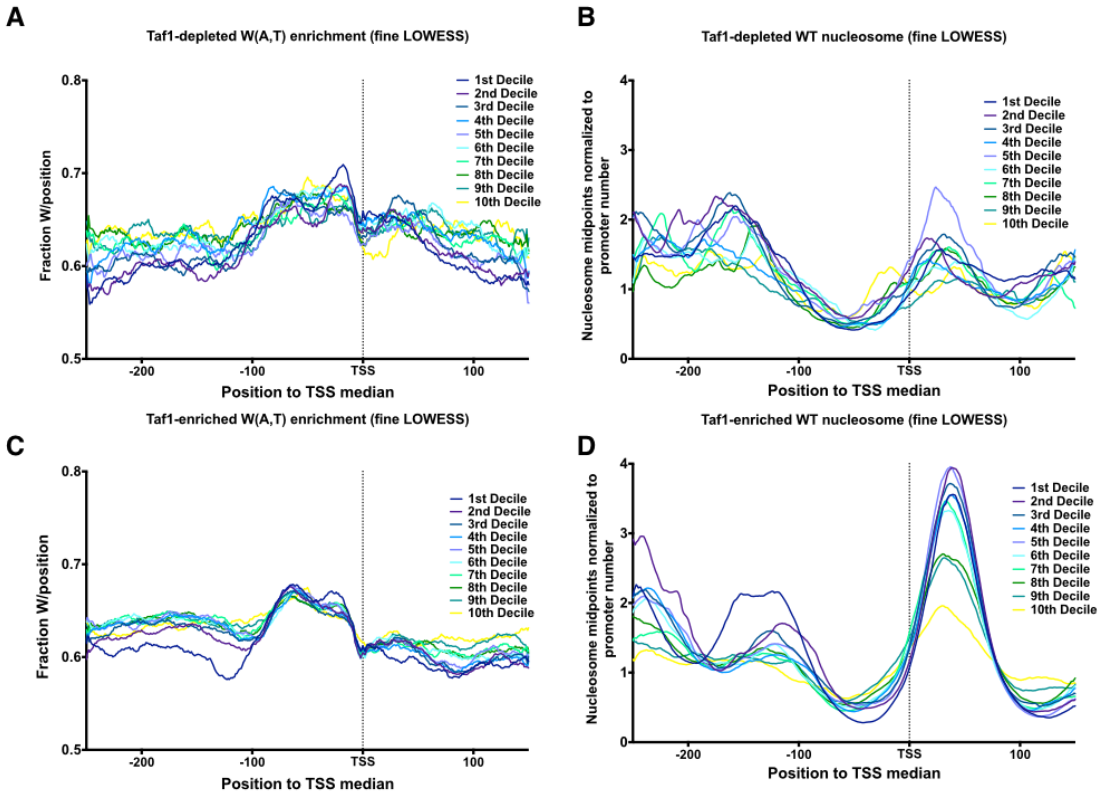


Figure 4-42. W (A/T) enrichment in promoter region does not correlate with expression level.

Aggregation of W (A/T) normalized to number of promoters in each expression decile (expression decile is determined by ranking all promoters; 1st: most expressed, 10th least expressed) are shown for: **A.** Taf1-depleted promoters. **C.** Taf1-enriched promoters. Nucleosome midpoints aggregation normalized to number of promoters in each expression decile are shown for **B.** Taf1-depleted promoters. **D.** Taf1-enriched promoters. All curves are smoothed by LOWESS.

Correlation of GTF localization and +1 nucleosome positioning

We noticed that Ssl2 (TFIIH) localization was closely related with +1 nucleosome positioning in **Figure 4-41,** suggesting that promoters have stereotypical nucleosome architecture relative to presumptive PIC position (mapped Ssl2 signals) while TSSs have evolved to be positioned across the intervening sequence, and this may confer distinct properties on promoters. To illustrate this, we divided Taf1-enriched promoters (the promoter class that has more defined localization of +1 nucleosome) into quintiles dependent on +1 nucleosome distance to median TSS. **Figure 4-43** presents the localization of Ssl2 (a subunit of GTF TFIIH) and nucleosomes midpoints in promoters binned by median TSS to +1 nucleosome distance quintiles in median TSS centered window. We observed a correlation between positioning of Ssl2 (TFIIH) and +1 nucleosome positioning in Taf1-enriched promoters as well as occupancies of Ssl2 and +1 nucleosome (**Figure 4-32**), suggesting they may have physical interaction/connection. Positioning of PIC complex may directly influence +1 nucleosome positioning during the initiation or *vice versa*. Highly expressed promoters are likely to have higher PIC occupancies, the correlation of PIC and +1 nucleosome positioning is consistent with our observation in **Figure 4-42** D which show a positive correlation between +1 nucleosome positioning and expression level. This is opposite of prediction by a model that suggests inhibitory role of nucleosome in expression. Highly expressed promoters may have more stereotypical chromatin structure and tighter regulation of nucleosome positioning. Further studies on relationship of PIC assembly and nucleosome will be interesting, and

any model of nucleosome influence in initiation should take this relationship into the consideration.

Nucleosome positioning is shifted in slow Pol II mutants

We observed Pol II activity mutants shift TSS distribution genome wide (**Figure 4-17**, **Figure 4-18**). During Pol II scanning, it is likely that Ssl2 is the factor that translocates Pol II to TSSs, thus directional shifts in TSS distribution may be translated to a shift in localization of Ssl2. Indeed, we detected the shifted Ssl2 localization in slow mutant H1085Y downstream by aggregation analysis (**Figure 4-44**). We predicted nucleosome positioning to be affected in this slow mutant because both initiation machinery and TSS selection were shifted. We examined nucleosome positioning in +1 nucleosome midpoint peak centered windows determined by the average of two WT (same digestion level, nucleosome data was highly correlated) +1 nucleosome midpoint peak datasets for each promoter. This was to detect changes in nucleosome positioning in mutants compared to the WT nucleosome positioning. +1 nucleosome peak calling was done using triweight kernel smoothening of the nucleosome midpoints mapped in median TSS aligned window and peak calling the local maximum over the smoothed midpoint profile.

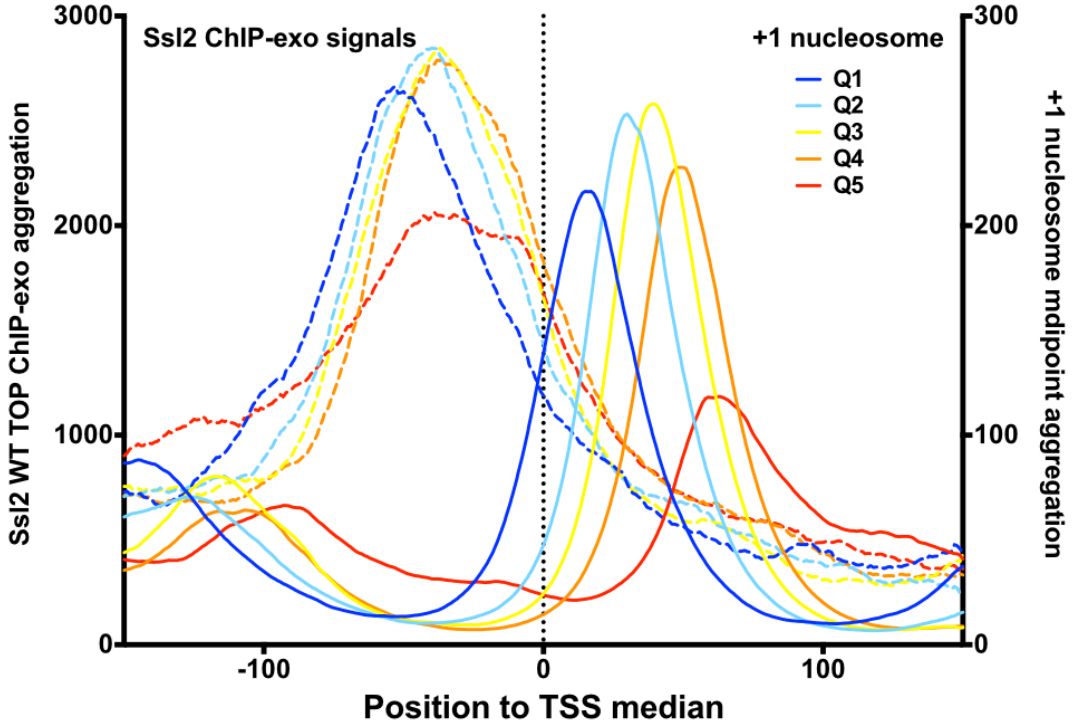


Figure 4-43. Ssl2 ChIP-exo signals and +1 nucleosome midpoints show positional correlation in median TSS centered promoter windows.

Taf1-enriched promoters >10 rpm expression are binned into quintile dependent on their distances between median TSS to +1 nucleosome peak. Aggregation of normalized Ssl2-ChIP-exo densities in each quintile are plotted in dotted lines on left *y* axis; +1 nucleosome midpoints aggregation in each quintile are plotted on right *y* axis.

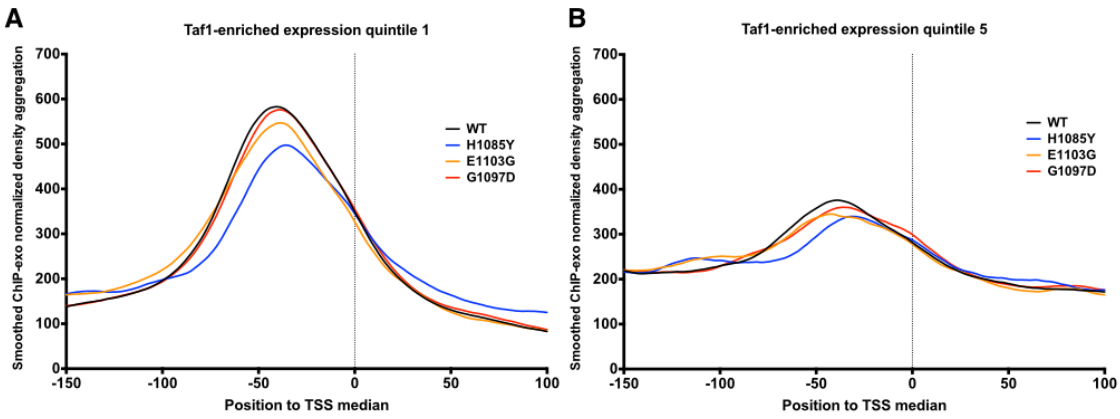


Figure 4-44. Ssl2 ChIP-exo signals in WT and Pol II activity mutants in median TSS aligned window show shift in slow mutant H1085Y.
Aggregation of smoothed normalized Ssl2 ChIP-exo TOP densities in WT and Pol II activity mutants for Taf1-enriched promoters in **A.** expression quintile 1 (most highly expressed) and **B.** expression quintle 5 (most lowly expressed).

We detected shifts of +1 nucleosome in the slow Pol II mutant *rpb1* H1085Y in the same direction as TSS distribution shift (downstream), and shifts were greater for +2 − +3 nucleosomes (**Figure 4-45**). The fast Pol II mutant *rpb1* E1103G did not seem to alter nucleosome positioning, or it was not detected in our current comparison method. Furthermore, the shift of nucleosome positioning in the slow Pol II mutant was sensitive to expression level as expected for a transcription-dependent effect and not an indirect effect on chromatin structure (**Figure 4-46,** expression quintile was determined based on all promoters).

Although the shift of +1- +3 we observed from the current comparison method was somewhat moderate, it is possible that a subpopulation of promoters have greater shifts of +1 nucleosomes. Shift in nucleosome positioning have been observed in nucleosome remodeler mutants and temperature sensitive *rpb1-1* mutant under high temperature previously (Alcid and Tsukiyama, 2014; Ganguli et al., 2014; van Bakel et al., 2013; Weiner et al., 2010). However, this is first time Pol II activity has been linked to modulation of nucleosome positioning. Transcription activity-dependent modulation of nucleosome positioning is intriguing, and we plan to verify these observations and expand our investigation on nucleosome positioning and relationship with PIC assembly in a more quantitative manner. We are currently awaiting sequencing on additional MNase-seq DNA libraries corresponding for Pol II mutants specifically matched to MNase digestion levels apparent for the WT samples. We hope to employ a more quantitative and complex methods for further analysis including determination of fragile

nucleosome and a measurement of "fragility" to verify if the mutant effects in fragile "0" nucleosomes (**Figure 4-45**) are reproducible.

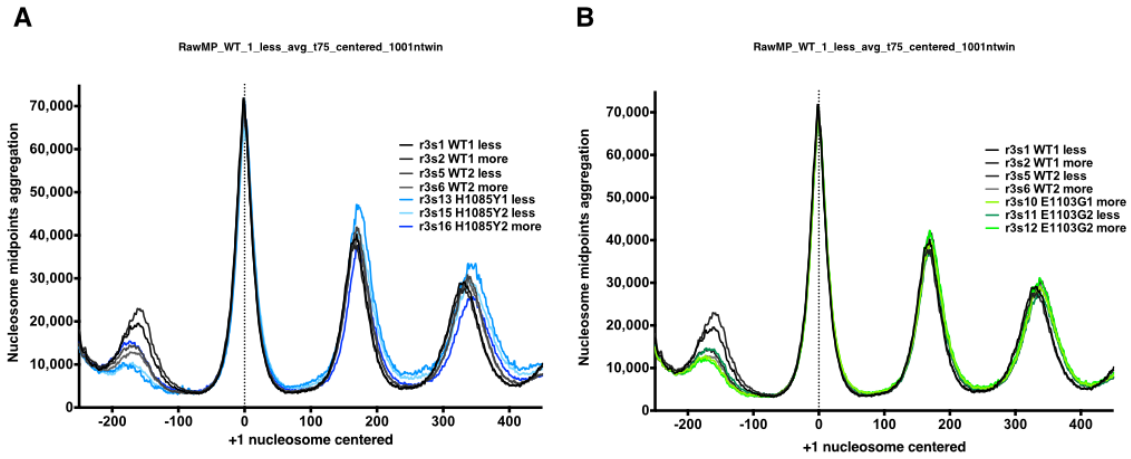


Figure 4-45. Nucleosome positioning in WT and Pol II activity mutants. Nucleosome midpoints aggregation in Taf1-enriched promoters in +1 nucleosome peak centered window for: **A.** WT and slow Pol II mutant *rpb1* H1085Y and **B.** WT and fast Pol II mutant *rpb1* E1103G libraries.

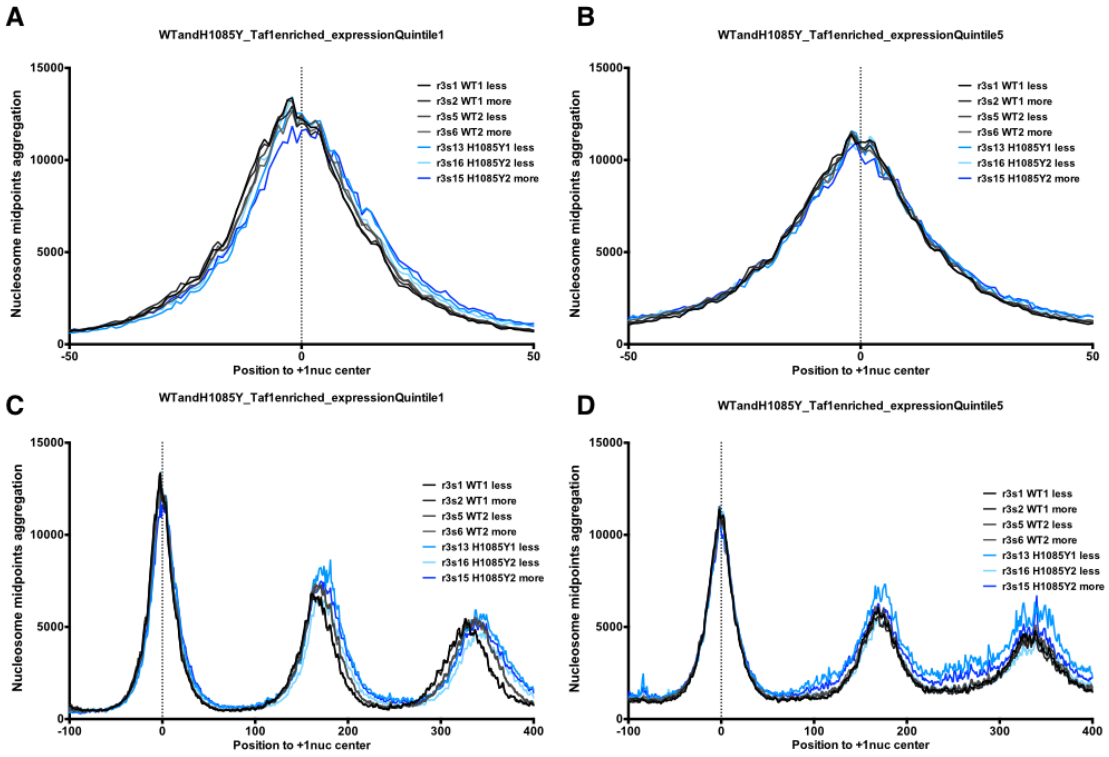


Figure 4-46. Slow Pol II mutant H1085Y shows directional shifts in +1 − +3 nucleosome positioning.

Nucleosome midpoint aggregation in Taf1-enriched promoters binned into expression quintile (expression quintile is determined by ranking all promoters) are shown in +1 nucleosome peak aligned window for WT and H1085Y libraries. **A.** −50 − +50 nt region in WT +1 nucleosome peak centered window for Taf1-enriched promoters in expression quintile 1 (most highly expressed). **B.** −50 − +50 nt region in WT +1 nucleosome peak centered window for Taf1-enriched promoters in expression quintile 5 (most lowly expressed). **C.** Same as A, but in −100 − +400 nt region. **D.** Same as B, but in −100 − +400 nt region.

Is scanning mechanism for TSS selection limited to *S. cerevisiae*?

A powerful model for TSS selection in *S. cerevisiae* proposed by Giardina and Lis that can explain various existing data is that Pol II is recruited to core promoter element and scans directionally downstream for TSS (Giardina and Lis, 1993). Key evidence for the model presented by Kuehner and Brow were discussed in Chapter I (Kuehner and Brow, 2006). Furthermore, the localization of all components of the PIC upstream of TSSs (Rhee and Pugh, 2012) indicates that the PIC components must move to sites of TSSs for current models of initiation to be possible. Recent studies have shown that other eukaryotes, including vertebrates, utilize multiple TSS at a large fraction of their promoters and show dispersed distribution of TSSs similar to *S. cerevisiae* (Chen et al., 2013; Choi et al., 2002; Consortium et al., 2014; Haberle et al., 2014; Hoskins et al., 2011; Li et al., 2015). TSS analyses in zebrafish at different growth stages revealed that gene expression in oocytes (maternal expression) and zygotes (zygotic expression) is driven by two distinct types of promoters (Haberle et al., 2014). Maternal TSSa have WW (A/T rich) boxes at ~−30 nt positions associated with individual TSSs, while zygotic TSSs for the same genes (only short distances away from maternal TSSs) had no individually associated promoter motif found. We hypothesized that zygotic promoters utilized scanning mechanism similarly to *S. cerevisiae*. Scanning Pol II (for example, *S. cerevisiae* and possibly zygotic promoters in zebrafish and in other eukaryotes) in the promoter region can potentially "see" all $Y_{-1}R_{+1}$ motifs within the TSS region because the region will be scanned. TSS selection therefore will heavily rely on usable YRs in the region, and the density for TSSs will correlate with the density

of available $Y_{-1}R_{+1}$ motifs (**Figure 4-47** A). TSSs that are individually associated with sequence elements at a certain distance upstream (for example, maternal promoters in zebrafish) will not be so dependent on $Y_{-1}R_{+1}$ motifs density in the region because Pol II will not "see" all the YR motifs, because it will be directed to particular motifs by the upstream elements (**Figure 4-47** B).

We measured the YR motif availability within regions of TSS "spread" for zebrafish promoters (overall sets of maternal and zygotic promoters). Specifically, we examined used YR motifs (usage >0) and those used above a 2% threshold (usage > 2% of overall density in promoter). Our model (**Figure 4-47**) predicts that the numbers of YR elements used relative to those available within the TSS spread region will be greater for zygotic promoters than for maternal promoters. Greater versus lesser YR usage (given similar YR availability) also will mean that dispersion of starts will be different even when normalizing for TSS regions of the same width. In other words, lower usage will result in spikier TSS distribution. Intriguingly, zygotic zebrafish promoters showed higher YR motif usage than maternal promoters, although the YR availability was similar to maternal promoters. In fact, the usage of YR in zygotic zebrafish promoters appeared very similar to that in *S. cerevisiae* (**Figure 4-48**). In addition, maternal zebrafish TSSs showed spikier TSS distribution than zygotic zebrafish TSSs as discussed above (lower YR site usage than zygotic zebrafish; however similar YR reads density, **Figure 4-49**). This observation is consistent with Pol II scanning for TSSs at zebrafish zygotic promoters.

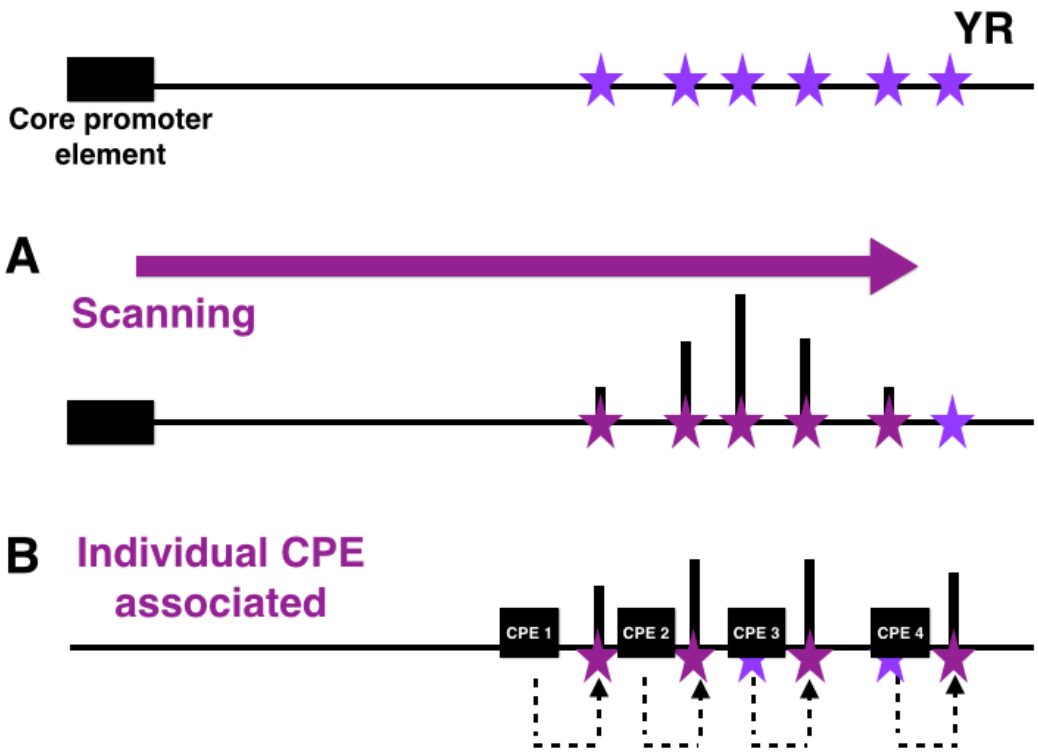


Figure 4-47. $Y_{-1}R_{+1}$ motif usage may differentiate TSSs selected by scanning from TSSs individually associated with core promoter elements.

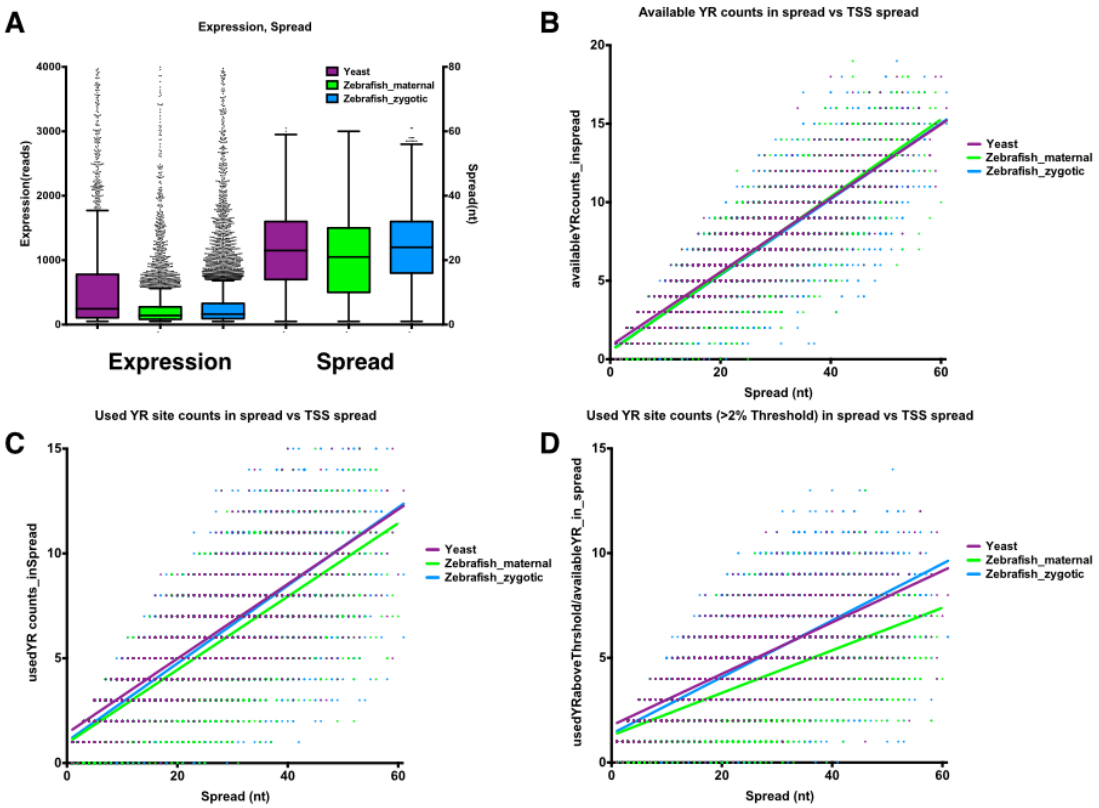


Figure 4-48. $Y_{-1}R_{+1}$ availability and usage for yeast and zebrafish promoters. **A.** Expression (read count) and TSS spread for 61 nt promoter region centered in median TSS are shown for representative yeast WT (library 446), zebrafish maternal (512 cells) and zebrafish zygotic (Prim 20 cells) TSS libraries. **B.** Number of available $Y_{-1}R_{+1}$ sites (*y* axis) in promoters that have each TSS spread (*x* axis) are shown as scatter plot for three libraries. Lines represent the linear regression. **C.** Same as B for number of used YR sites (usage>0). D. Same as B for number of used YR sites above threshold (usage> 2% of total reads in promoter).

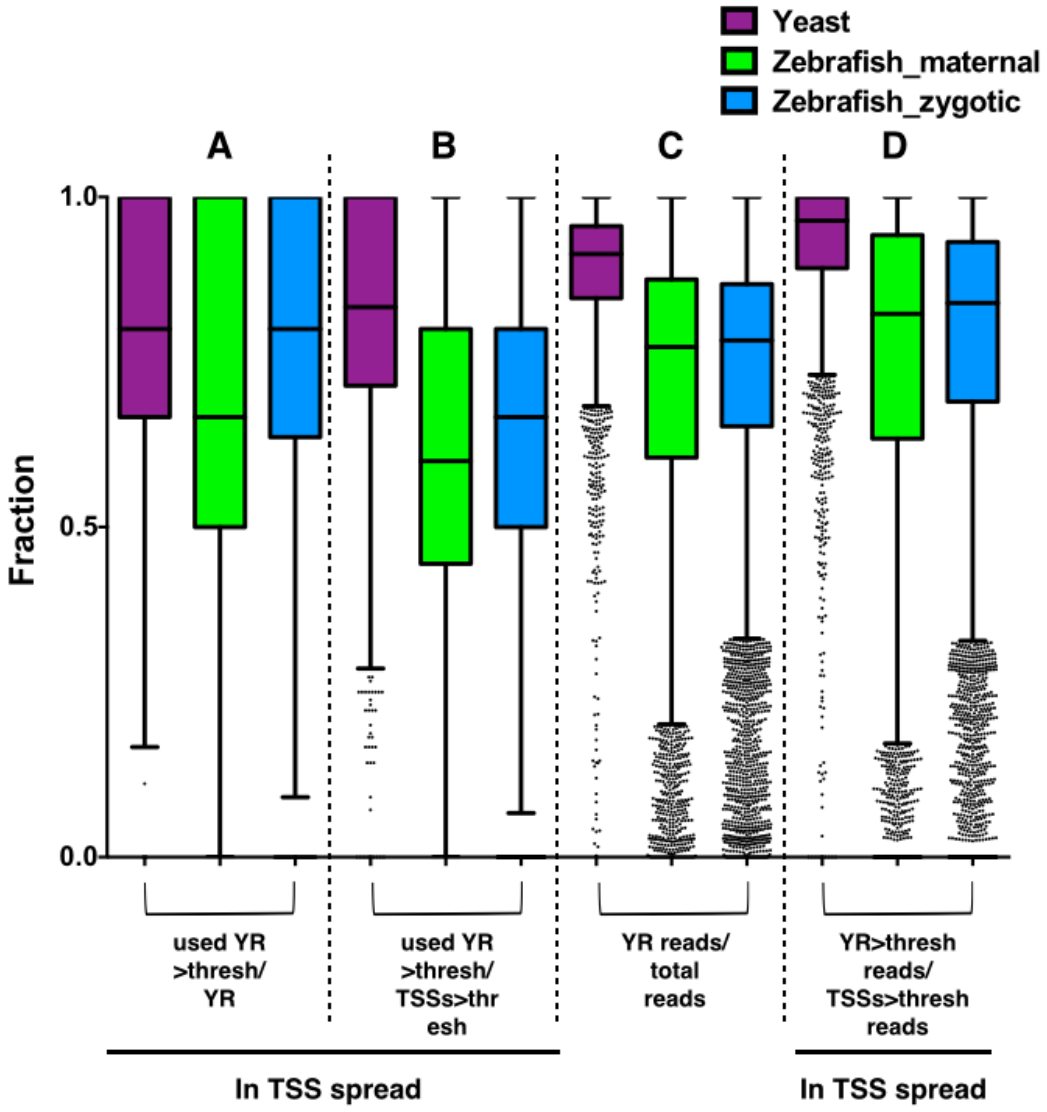


Figure 4-49. $Y_{-1}R_{+1}$ usage/availability fraction in yeast and zebrafish promoters. **A.** YR sites within TSS spread used above threshold (usage> 2% of total reads in promoter) fraction in available YR sites in TSS spread. **B.** YR sites within spread used above threshold fraction in TSS sites (YR and non-YR) within TSS spread used above threshold. **C.** YR site reads fraction in total reads in promoter. **D.** YR sites within TSS spread used above threshold reads fraction in TSS sites (YR and non-YR) within TSS spread used above threshold reads.

Discussion

Understanding initiation mechanisms in eukaryotes is essential for understanding gene expression, because initiation is a major regulatory hub. TSS selection/utilization defines the 5′ of the transcripts, and generates transcripts with different lengths of 5′ UTR. This process may have impacts on downstream RNA processing activities and regulation including translation efficiency. Genome wide TSS mapping by deep sequencing has facilitated our understanding of initiation regulation, however, the field was lacking a genome wide study of TSSs and their relationships to promoter architecture. Moreover, a mechanistic investigation on the influence of Pol II and other key initiation factors' activities genome wide was a needed effort.

We utilized SOLiD 5′-RNA sequencing to map TSS genome wide in WT and Pol II activity/GTF mutants and conducted several analyses to investigate promoter architecture regulation of initiation using TSS selection as a measurable aspect of the initiation process. We discovered genome wide polar effects on TSS utilization in Pol II activity and GTF mutants for all promoter classes, consistent with their alteration of TSS distributions at TATA-containing model genes. This strongly supported directional Pol II scanning mechanism for *S. cerevisiae* at most if not all promoters. We proposed two models for the reason Pol II activity/GTF mutants shift distributions of TSS usage directionally. First, Pol II activity/GTF mutants alter TSS motif sequence specificity, leading to TSS distribution shift. Second, the observation of TSS motifs usage alteration is the resultant of shifting initiation efficiency window by Pol II activity/GTF mutants. Both models fit with existing data, therefore a single promoter based quantitative study

will be required to further investigate the polar effects on TSS selection conferred by Pol II activity/GTF mutants.

Our data suggests a model where promoter architecture is a determinant of initiation properties, these properties can be altered by initiation factor mutants, and thus promoters may be differentially sensitive to factor mutants based on their specific architectures. Pol II activity/GTF mutants that shift TSS usage distribution genome wide confer different sensitivities of shifting dependent on promoter architecture especially in TATA-containing promoters. Our results are consistent with a model that TATA element poses a constraint in TSS selection in terms of determining the efficiency window for scanning effectively.

Surprisingly, we discovered that Pol II activity/GTF mutants show a certain gene expression changes dependent on mutant class (fast or slow) correlating with the distance between core promoters and TSSs for both promoter classes. This observation is consistent with a model that promoters establish a certain level of transcription output by placing TSSs at specific distances. We hypothesize that alteration of Pol II/initiation activity changes the initiation efficiency window, breaking the pre-formed relationship between initiation properties and scanning distance by Pol II, leading to changes in promoter output of transcription. A single-promoter based future experiment with careful alteration of scanning distances with minimum change of sequence composition will be required to test this model. A very recent publication by Segal group (Lubliner et al., 2015) presented an experiment that utilized chimera constructs with two TATA-containing promoter regions each fused to a fluorescent protein YFP ORF and then

varied TATA element positions for the two promoters. A total of four sets of experiments were presented– comprising distance variance constructs of two TATA elements in two promoters, and sequence variants in other promoter attributes. Expression level (by fluorescence-activated cell sorting (FACS)) and TSSs positions (by 5′ RACE (rapid amplification of cDNA ends)) for ~5000 total constructs were measured separately. Distance variance constructs that placed TATA element ~−100 nt~ −50 nt relative to TSSs demonstrated that distance between TATA element and TSSs affected expression level (consistent with our results discussed above). Moreover, the optimal distance that supports maximum expression level at a particular promoter seemed to depend on core promoter element sequences. Changes in TATA element positions also altered TSS selection, consistent with our and others' data. These results are supportive of our results presented above, and we are currently seeking ways to utilize the published raw sequencing data to conduct a few more analyses.

To further understand the responsible core promoter elements for TSS selection in Taf1-enriched promoters, we selected 12 Taf1-enriched promoters to fuse to a reporter ORF and mutated proposed core promoter elements TATA-like and GAE (Rhee and Pugh, 2012; Seizl et al., 2011). Deletion/mutations at TATA-like or GAE in our candidate promoters show different degrees of decrease in expression level, in summary two promoters seem to be driven by GA element, one promoter by TATA-like element, and seven promoters did not show decrease in expression in either of elements. A very recent publication from the Segal group (Lubliner et al., 2015) presented their studies on GA 5-mer elements. They knocked out GAEs for 122 promoters and detected

insignificant changes in expression level except for two. The distances between these GAE and TSSs are not shown in the study, we'd like to utilize our TSS mapping to identify what fraction of tested GAEs are in presumptively optimal distances from TSSs that showed the enrichment peak in **Figure 4-36**.

We utilized our high resolution TSS data to test an existing model for TSS selection in different promoter classes previously proposed by Pugh group (Rhee and Pugh, 2012). Rhee and Pugh's model suggested that different promoter classes vary in TSS selection by distinctive participation of nucleosome positioning in PIC assembly. We examined our data to test three simple predictions of this model and our analyses were not supportive of the model.

In contrast to the view that nucleosomes act as barriers for transcription, we discovered a more dynamic and possibly passive way of nucleosome positioning's influence on initiation. We showed that assembly of PIC and +1 nucleosome positioning are closely correlated. Which is the dominant force is not yet known, however, testing of any model of how +1 nucleosome influences initiation process needs to bear in mind the relationship of PIC and +1 nucleosome in positioning. We observed the downstream shift in +1 ~ +3 nucleosome positioning in our slow Pol II mutant, the direction of shift interestingly same as shift in TSS distribution in this mutant. We also observed that downstream shift in +2 ~ +3 nucleosomes in our slow Pol II mutant greater in magnitude. We also showed that the directional shift of nucleosome positioning is sensitive to gene expression level, suggesting shifts are transcription dependent. Positioning of downstream nucleosomes have been suggested to be influenced by +1 nucleosome by

statistical positioning (steric repulsion) (Kaplan et al., 2009; Mavrich et al., 2008), thus it is possible the greater shifts we observe in +2 ~ +3 nucleosomes are the cascade effect caused by +1 nucleosome. We cannot rule out the shifts we observe in nucleosome positioning are mainly due to defective elongation by slow Pol II mutant, and this will be tested by examining nucleosome positioning in the GTF TFIIB mutant *sua7-1* in which all the primary defects arise during initiation.

We observed influence of promoter architecture on initiation output changes conferred by Pol II activity/GTF mutants. Therefore, we propose and summarize the promoter architectural determinants of TSS selection/utilization as follows (**Figure 4-50**). 1. The core promoter element is a key identifier for distinctive promoter architectures, different core promoter elements show distinct dependencies of initiation factors (for example Taf1). 2. Sequence context in the promoter region between core promoter element and TSSs (scanning region) favors or repels nucleosome positioning in the region. We observed T enrichment in this region most distinctive in highly expressed promoters in both promoter class, and hypothesized T enrichment favors Pol II scanning thermodynamically. Supportive evidence are presented by the study by the Segal group (Lubliner et al., 2015), in which results suggested T/C enrichment in this region increased the output of promoter within an otherwise identical TSSs region sequence context. 3. Distance between core promoter element and TSSs have been shown to influence various initiation properties including sensitivity of TSS usage shifts, expression level/noise by initiation modulators. We proposed a relationship between initiation efficiency window and TSS selection to explain our observations. 4. +1

nucleosome positioning and occupancy. We revealed relationship between PIC assembly and +1 nucleosome positioning was much more stereotypical than TSS positioning relative to +1 nucleosome positioning, and suggested that nucleosome positioning may play a role more than as a barrier for transcription. We are curious why nucleosome positioning is shifted in slow Pol II mutant in a manner sensitive to expression level, and are in process of expanding our investigation on nucleosomes effect on TSS selection and other initiation regulations.

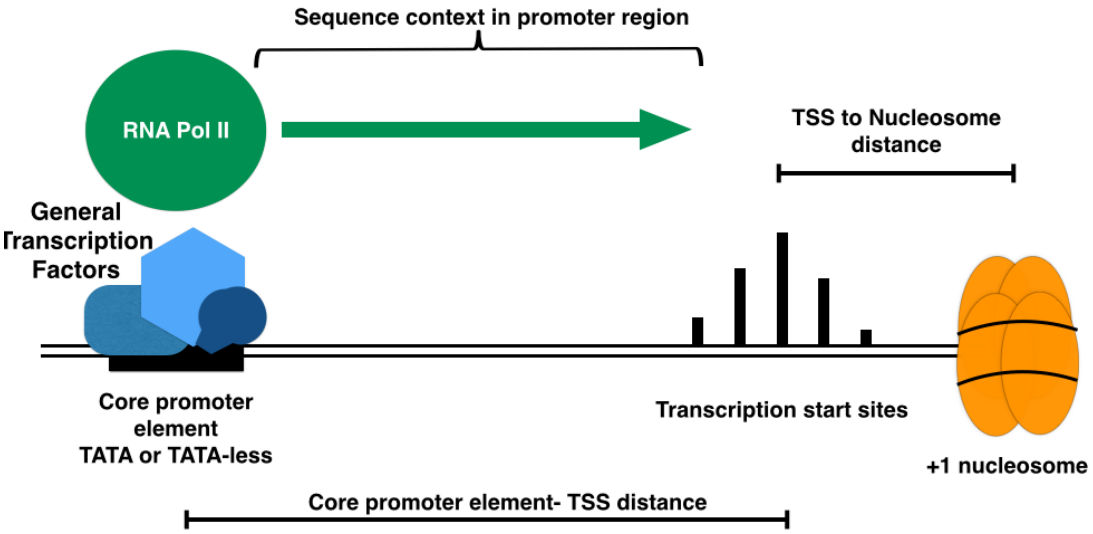


Figure 4-50. Promoter architectural determinants for TSS selection in *S. cerevisiae*.

Finally, we present analyses that zebrafish TSSs share similar distribution characteristics to *S. cerevisiae* consistent with a conserved mechanism for their specification, *i.e.* promoter scanning. We used a bioinformatics analysis that calculates usage of initiator motif YR versus the available YR and are able to differentiate TSS selection by scanning Pol II and by individually associated core promoters. My preliminary analysis on a set of human promoters (~1000) also suggests that human TSSs have similar distribution characteristics as yeast and zebrafish zygotic promoters. Human promoters are highly enriched in CpG, which is a distinct characteristic of mammalian promoters, and the YR usage analysis numbers are affected by this CpG bias. However, by analyzing different YR usages across individual YR motifs, we found that CpGs were disfavored for initiation. When we examine non-CpG YR usage relative to non-CpG YR density we found that human TSS distribution was distinguishable from zebrafish maternal promoters and more similar to yeast or zebrafish zygotic promoters. We are aiming to expand our human promoter analysis by taking a recent published data from the Lis lab (Core et al., 2014) and analyze more promoters with greater coverage depths for TSSs reads. We also plan to perform a similar analysis with very recent *S. pombe* TSS sequencing (Li et al., 2015). We predict the promoters with dispersed multiple TSS usage (~50% of *S. pombe* promoters) will show similar behavior as *S. cerevisiae* and zygotic zebrafish in YR usage.

More recent studies are uncovering that other eukaryotes show many similarities on TSS selection with *S. cerevisiae*. Our findings on TSS selection and its control of promoter architecture may shed light on mechanisms of TSS selection in other

eukaryotes. We would like to expand our studies on functionality and impact on transcription output of transcripts with various TSSs. The lengths of 5′ UTR (untranslated region) has been linked to translation efficiency (Arribere and Gilbert, 2013; McManus et al., 2014), thus linking transcription start site selection and translation. My preliminary analysis showed that 51 of 115 uORFs (translated ORFs upstream of stereotypical ORFs) (Ingolia et al., 2009) in our promoter list are within our measured TSS spreads. TSSs that are upstream of these 51 uORFs may generate transcripts functionally different from other transcripts influencing gene expression level and expression noise. Pol II activity/GTF mutants that alter TSS distribution may alter uORFs transcription, we plan to conduct the uORFs analyses using data published by Jacquier group in comparison of our TSS sequencing (Malabat et al., 2015).

CHAPTER V

SUMMARY AND FUTURE DIRECTIONS

Summary

The work presented in this dissertation explores mechanisms of how Pol II catalytic activity influences initiation processes, more specifically transcription start site (TSS) selection/utilization *in vivo* in *Saccharomyces cerevisiae* (budding yeast). We explored influence of Pol II activity on TSS selection and other transcriptional controls *in vivo* utilizing genetic and biochemical approach and expanded our investigation genome wide by deep sequencing and bioinformatics analyses.

We observed that Pol II active site mutants that altered catalytic activities *in vitro* conferred a polar effect on distribution of TSS *in vivo* by analyzing several model genes. The magnitude of shifts correlated tightly with extent of activity alterations (Chapter II, III). I investigated whether conditional growth phenotypes that correlate with different transcriptional defects also correlated with TSS selection changes. I found that the MPA^s growth phenotype was closely correlated with the type of upstream TSS shifting defects widely conferred by Pol II fast mutants and initiation factor TFIIF mutants (Chapter II, III, (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012)). I contributed substantially to an investigation utilizing high throughput quantitative analysis pE-MAP that explored genetic interactions among thousands of factors and Pol II alleles in collaboration with Krogan lab (second subchapter in Chapter II). I showed that pE-MAP

is able to cluster mutants by their genetic phenotypes and interactions with other factors. I discovered novel functions for Sub1: *sub1*Δ altered TSS utilization on its own and also modified TSS defects of Pol II slow mutants but is bypassed by Pol II fast mutants. The exacerbation, suppression, or epistasis of Pol II mutant TSS defects by *sub1*Δ are mirrored in its genetic interactions with Pol II mutants, which are unique among Pol II mutant genetic interactors (Chapter II, III). I furthered my study of Pol II activity mutants combined with GTF and other transcription factor mutants to ask if TSS utilization defects might be causative of general and conditional growth defects observed in Pol II activity mutants. I concluded that although some conditional growth defects were correlated with (and possibly caused by) TSS defects, TSS defects do not appear to directly cause general growth defects unless the initiation defects pass a certain severity in which case cell growth becomes severely impaired (Chapter III).

To mechanistically explore TSS utilization and their alterations in Pol II activity mutants in a global scale, I utilized three types of genome wide sequencing, and mapped TSSs, GTF occupancies, and nucleosomes in *S. cerevisiae*. I conducted the analyses with the genome wide data bearing seven main questions in mind. 1. How are promoters of different classes evolutionary different? 2. Do Pol II activity/GTF mutants influence TSS selection in different promoter classes differently? 3. Can Pol II activity/GTF mutants allow us to probe for functional differences in promoter architecture? 4. How do Pol II activity/GTF mutants shift TSS? 5. What are the core promoter elements for TATA-less promoters, a promoter class containing ~80% of eukaryotic promoters? 6. How does the

+1 nucleosome contribute to shaping initiation properties such as TSS selection or PIC assembly? 7. Do other eukaryotic Pol IIs scan for TSS similar to *S. cerevisiae*?

Future directions

There are various future studies planned for investigation into the mechanism of TSS selection/utilization and alteration by Pol II activity/GTF mutants. We aim to further our investigation to understand the questions described above as well as the significance of having multiple TSSs in each promoter and their regulation on usage. Do multiple TSSs in each promoter lead to transcripts with different expression profiling and functionality?

I list the relatively near future studies we are interested in as follows. 1. Compare our TSS mapping with recently published data by others by taking mapped reads in others' sequencing data (Malabat et al., 2015) and map the reads in our promoter window to compare with our data by a position-based correlation used in our study. 2. We would like to investigate if promoter architecture, more specifically the distance between the core promoter element and the TSS (scanning distance), contributes to expression levels and expression noise. We will utilize single promoter-based quantitative experiments and alter scanning distances with minimum changes in sequence composition. A recent publication by the Segal group has performed a set of experiments similar in concept (Lubliner et al., 2015); however the measurement of TSSs and transcription activity are somewhat disconnected in this study (expression of fluorescent protein by fluorescence cytometry was used as measure of transcription

activity) . We will choose several representative promoters with medium-high expression levels in Taf1-depleted (the majority of promoters in this class are TATA-containing) and Taf1-enriched (mostly TATA-less) promoter classes and utilize spike-in controls to more accurately and quantitatively measure the transcript level as measurement of expression. 3. We will also continue our research on the role of nucleosome positioning in initiation regulation. We are currently awaiting sequencing of more libraries of the mutants used in Chapter IV to verify our findings. We hope to statistically assess mutant effect on nucleosome positioning. We will ask if the +1 ~ +3 nucleosome positioning shifts by Pol II slow mutant H1085Y are mainly caused by initiation defect rather than elongation defects. For this matter, we will utilize a GTF mutant *sua7-1* that shows TSS shifts in the same direction and in a similar degree as H1085Y, however the main defects of the mutant are presumably in initiation and not elongation. 4. We plan to utilize our bioinformatics analyses on initiator motifs usage in to investigate other eukaryotes to expand our analyses on TSS selection mechanism. My preliminary analysis with relatively small set of human promoters (~1000) is consistent with our hypothesis that non-TATA human promoters are scanned during TSS selection. We will utilize more recent and extensive human TSS data by the Lis group (Core et al., 2014) and analyze a greater number of human promoters with the same bioinformatics pipeline. We also plan to utilize very recent publication on *S. pombe* TSS mapping for similar analyses (Li et al., 2015) as well as for Drosophila promoters (Hoskins et al., 2011).

REFERENCES

Aguilar, P.S., Frohlich, F., Rehman, M., Shales, M., Ulitsky, I., Olivera-Couto, A., Braberg, H., Shamir, R., Walter, P., Mann, M.*, et al.* (2010). A plasma-membrane E-MAP reveals links of the eisosome with sphingolipid metabolism and endosomal trafficking. Nat Struct Mol Biol *17*, 901-908.

Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol Cell *46*, 115-124.

Aitken, S., Alexander, R.D., and Beggs, J.D. (2011). Modelling reveals kinetic advantages of co-transcriptional splicing. PLoS Comput Biol *7*, e1002215.

Alcid, E.A., and Tsukiyama, T. (2014). ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape. Genes Dev *28*, 2348-2360.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Mol Cell *40*, 582-593.

Amberg, D.C., Burke, D., Strathern, J.N., Burke, D., and Cold Spring Harbor Laboratory. (2005). Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual, 2005 edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Apone, L.M., Virbasius, C.A., Holstege, F.C., Wang, J., Young, R.A., and Green, M.R. (1998). Broad, but not universal, transcriptional requirement for yTAFII17, a histone H3-like TAFII present in TFIID and SAGA. Mol Cell *2*, 653-661.

Archambault, J., and Friesen, J.D. (1993). Genetics of eukaryotic RNA polymerases I, II, and III. Microbiol Rev *57*, 703-724.

Archambault, J., Jansma, D.B., Kawasoe, J.H., Arndt, K.T., Greenblatt, J., and Friesen, J.D. (1998). Stimulation of transcription by mutations affecting conserved regions of RNA polymerase II. J Bacteriol *180*, 2590-2598.

Archambault, J., Lacroute, F., Ruet, A., and Friesen, J.D. (1992). Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. Mol Cell Biol *12*, 4142-4152.

Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Res *23*, 977-987.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. Cell *120*, 183-193.

Barberis, A., Muller, C.W., Harrison, S.C., and Ptashne, M. (1993). Delineation of two functional regions of transcription factor TFIIB. Proc Natl Acad Sci U S A *90*, 5628-5632.

Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell *116*, 699-709.

Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chen, S., Mann, M.*, et al.* (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. Cell *152*, 909-922.

Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C., Kulbachinskiy, A., and Murakami, K.S. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. J Biol Chem *289*, 24549-24559.

Bednar, J., Studitsky, V.M., Grigoryev, S.A., Felsenfeld, G., and Woodcock, C.L. (1999). The nature of the nucleosomal barrier to transcription: direct observation of paused intermediates by electron cryomicroscopy. Mol Cell *4*, 377-386.

Beltrao, P., Cagney, G., and Krogan, N.J. (2010). Quantitative genetic interactions reveal biological modularity. Cell *141*, 739-745.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) *57*, 289-300.

Benschop, J.J., Brabers, N., van Leenen, D., Bakker, L.V., van Deutekom, H.W., van Berkum, N.L., Apweiler, E., Lijnzaad, P., Holstege, F.C., and Kemmeren, P. (2010). A consensus of core protein complex compositions for Saccharomyces cerevisiae. Mol Cell *38*, 916-928.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol *17*, 251-256.

Bergkessel, M., Whitworth, G.B., and Guthrie, C. (2011). Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast. RNA *17*, 1461-1478.

Berroteran, R.W., Ware, D.E., and Hampsey, M. (1994). The sua8 suppressors of Saccharomyces cerevisiae encode replacements of conserved residues within the largest

subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations. Mol Cell Biol *14*, 226-237.

Boeke, J.D., Trueheart, J., Natsoulis, G., and Fink, G.R. (1987). 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol *154*, 164-175.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Bondarenko, V.A., Steele, L.M., Ujvari, A., Gaykalova, D.A., Kulaeva, O.I., Polikanov, Y.S., Luse, D.S., and Studitsky, V.M. (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. Mol Cell *24*, 469-479.

Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F.*, et al.* (2013). From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell *154*, 775-788.

Brogaard, K., Xi, L., Wang, J.P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. Nature *486*, 496-501.

Brueckner, F., and Cramer, P. (2008). Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. Nat Struct Mol Biol *15*, 811-818.

Bucheli, M.E., and Buratowski, S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. EMBO J *24*, 2150-2160.

Bucheli, M.E., He, X., Kaplan, C.D., Moore, C.L., and Buratowski, S. (2007). Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. RNA *13*, 1756-1764.

Buratowski, S., Sopta, M., Greenblatt, J., and Sharp, P.A. (1991). RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. Proc Natl Acad Sci U S A *88*, 7509-7513.

Buratowski, S., and Zhou, H. (1993). Functional domains of transcription factor TFIIB. Proc Natl Acad Sci U S A *90*, 5633-5637.

Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O.*, et al.* (2008). eSGA: E. coli synthetic genetic array analysis. Nat Methods *5*, 789-795.

Cabart, P., Jin, H., Li, L., and Kaplan, C.D. (2014). Activation and reactivation of the RNA polymerase II trigger loop for intrinsic RNA cleavage and catalysis. Transcription *5*.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol Cell *40*, 571-581.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A., Gotte, M., Konigsberg, W., and Cameron, C.E. (2009). Nucleic acid polymerases use a general acid for nucleotidyl transfer. Nat Struct Mol Biol *16*, 212-218.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. FEBS J *276*, 1494-1505.

Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E., and Crouch, R.J. (2003). Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. Mol Cell *11*, 807-815.

Chen, B.S., and Hampsey, M. (2004). Functional interaction between TFIIB and the Rpb2 subunit of RNA polymerase II: implications for the mechanism of transcription initiation. Mol Cell Biol *24*, 3983-3991.

Chen, H.T., and Hahn, S. (2004). Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. Cell *119*, 169-180.

Chen, H.T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat Struct Mol Biol *14*, 696-703.

Chen, R.A., Down, T.A., Stempor, P., Chen, Q.B., Egelhofer, T.A., Hillier, L.W., Jeffers, T.E., and Ahringer, J. (2013). The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. Genome Res *23*, 1339-1347.

Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J *4*, 3273-3280.

Cheung, A.C., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. EMBO J *30*, 4755-4763.

Cho, E.J., and Buratowski, S. (1999). Evidence that transcription factor IIB is required for a post-assembly step in transcription initiation. Journal of Biological Chemistry *274*, 25807-25813.

Choi, W.S., Yan, M., Nusinow, D., and Gralla, J.D. (2002). In vitro transcription and start site selection in Schizosaccharomyces pombe. J Mol Biol *319*, 1005-1013.

Clark, T.A., Sugnet, C.W., and Ares, M., Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science *296*, 907-910.

Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S., and Krogan, N.J. (2007a). Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol Cell Proteomics *6*, 439-450.

Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M.*, et al.* (2007b). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature *446*, 806-810.

Collins, S.R., Roguev, A., and Krogan, N.J. (2010). Quantitative genetic interaction mapping using the E-MAP approach. Methods Enzymol *470*, 205-231.

Collins, S.R., Schuldiner, M., Krogan, N.J., and Weissman, J.S. (2006). A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol *7*, R63.

Conesa, C., and Acker, J. (2010). Sub1/PC4 a chromatin associated protein with multiple functions in transcription. RNA Biol *7*, 287-290.

Consortium, F., the, R.P., and Clst (2014). A promoter-level mammalian expression atlas. Nature *507*, 462-470.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. Science *209*, 1406-1414.

Corden, J.L. (2008). Yeast Pol II start-site selection: the long and the short of it. EMBO Rep *9*, 1084-1086.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311-1320.

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

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J.*, et al.* (2008). Structure of eukaryotic RNA polymerases. Annu Rev Biophys *37*, 337-352.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863-1876.

Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci U S A *94*, 11456-11460.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Da, L.T., Wang, D., and Huang, X. (2012). Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. J Am Chem Soc *134*, 2399-2406.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci U S A *103*, 5320-5325.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525-532.

de la Mata, M., Munoz, M.J., Allo, M., Fededa, J.P., Schor, I.E., and Kornblihtt, A.R. (2011). RNA Polymerase II Elongation at the Crossroads of Transcription and Alternative Splicing. Genet Res Int *2011*, 309865.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680-686.

Desmoucelles, C., Pinson, B., Saint-Marc, C., and Daignan-Fornier, B. (2002). Screening the yeast "disruptome" for mutants affecting resistance to the immunosuppressive drug, mycophenolic acid. J Biol Chem *277*, 27036-27044.

Du, H.N., Fingerman, I.M., and Briggs, S.D. (2008). Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4. Genes Dev *22*, 2786-2798.

Dvir, A. (2002). Promoter escape by RNA polymerase II. Biochimica Et Biophysica Acta-Gene Structure and Expression *1577*, 208-223.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J Biol Chem *266*, 71-75.

Eichner, J., Chen, H.T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. EMBO J *29*, 706-716.

Exinger, F., and Lacroute, F. (1992). 6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae. Curr Genet *22*, 9-11.

Faitar, S.L., Brodie, S.A., and Ponticelli, A.S. (2001). Promoter-specific shifts in transcription initiation conferred by yeast TFIIB mutations are determined by the sequence in the immediate vicinity of the start sites. Mol Cell Biol *21*, 4427-4440.

Feaver, W.J., Svejstrup, J.Q., Bardwell, L., Bardwell, A.J., Buratowski, S., Gulyas, K.D., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1993). Dual roles of a multiprotein complex from S. cerevisiae in transcription and DNA repair. Cell *75*, 1379-1387.

Fiedler, D., Braberg, H., Mehta, M., Chechik, G., Cagney, G., Mukherjee, P., Silva, A.C., Shales, M., Collins, S.R., van Wageningen, S.*, et al.* (2009). Functional organization of the S. cerevisiae phosphorylation network. Cell *136*, 952-963.

Field, Y., Fondufe-Mittendorf, Y., Moore, I.K., Mieczkowski, P., Kaplan, N., Lubling, Y., Lieb, J.D., Widom, J., and Segal, E. (2009). Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. Nat Genet *41*, 438-445.

Fishburn, J., and Hahn, S. (2012). Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. Mol Cell Biol *32*, 12-25.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., and Reinberg, D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A *88*, 9999-10003.

Freire-Picos, M.A., Krishnamurthy, S., Sun, Z.W., and Hampsey, M. (2005). Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. Nucleic Acids Res *33*, 5045-5052.

Furter-Graves, E.M., and Hall, B.D. (1990). DNA sequence elements required for transcription initiation of the Schizosaccharomyces pombe ADH gene in Saccharomyces cerevisiae. Mol Gen Genet *223*, 407-416.

Furter-Graves, E.M., Hall, B.D., and Furter, R. (1994). Role of a small RNA pol II subunit in TATA to transcription start site spacing. Nucleic Acids Res *22*, 4932-4936.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397-405.

Ganguli, D., Chereji, R.V., Iben, J.R., Cole, H.A., and Clark, D.J. (2014). RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. Genome Res *24*, 1637-1649.

Garcia, A., Collin, A., and Calvo, O. (2012). Sub1 associates with Spt5 and influences RNA polymerase II transcription elongation rate. Mol Biol Cell *23*, 4297-4312.

Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., and Ponticelli, A.S. (2004). Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. Molecular and Cellular Biology *24*, 10975-10985.

Giardina, C., and Lis, J.T. (1993). DNA melting on yeast RNA polymerase II promoters. Science *261*, 759-762.

Giardina, C., Perez-Riba, M., and Lis, J.T. (1992). Promoter melting and TFIID complexes on Drosophila genes in vivo. Genes Dev *6*, 2190-2200.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell *45*, 814-825.

Ginsburg, D.S., Govind, C.K., and Hinnebusch, A.G. (2009). NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5. Mol Cell Biol *29*, 6473-6487.

Gleghorn, M.L., Davydova, E.K., Basu, R., Rothman-Denes, L.B., and Murakami, K.S. (2011). X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. Proc Natl Acad Sci U S A *108*, 3566-3571.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*, 1876-1882.

Goel, S., Krishnamurthy, S., and Hampsey, M. (2012). Mechanism of start site selection by RNA polymerase II: interplay between TFIIB and Ssl2/XPB helicase subunit of TFIIH. J Biol Chem *287*, 557-567.

Goldman, S.R., Sharp, J.S., Vvedenskaya, I.O., Livny, J., Dove, S.L., and Nickels, B.E. (2011). NanoRNAs prime transcription initiation in vivo. Mol Cell *42*, 817-825.

Green, M.R. (2000). TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. Trends Biochem Sci *25*, 59-63.

Greger, I.H., Aranda, A., and Proudfoot, N. (2000). Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *97*, 8415-8420.

Greger, I.H., and Proudfoot, N.J. (1998). Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. EMBO J *17*, 4771-4779.

Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D., and Hughes, T.R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. Mol Cell Biol *24*, 5534-5547.

Grosso, A.R., de Almeida, S.F., Braga, J., and Carmo-Fonseca, M. (2012). Dynamic transitions in RNA polymerase II density profiles during transcription termination. Genome Res *22*, 1447-1456.

Grunberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nat Struct Mol Biol *19*, 788-796.

Guthrie, C., and Fink, G.R. (2002). Guide to Yeast Genetics and Molecular and Cell Biology (Academic Press).

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M.*, et al.* (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381-385.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol *11*, 394-403.

Hahn, S., Hoar, E.T., and Guarente, L. (1985). Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *82*, 8562-8566.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.

Haight, F.A. (1967). Handbook of the Poisson distribution (New York,: Wiley).

Hampsey, M. (1998a). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiology and Molecular Biology Reviews *62*, 465-+.

Hampsey, M. (1998b). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev *62*, 465-503.

Hampsey, M. (2006). The Pol II initiation complex: finding a place to start. Nat Struct Mol Biol *13*, 564-566.

Handschumacher, R.E., and Pasternak, C.A. (1958). Inhibition of orotidylic acid decarboxylase, a primary site of carcinostasis by 6-azauracil. Biochim Biophys Acta *30*, 451-452.

Harel-Sharvit, L., Eldad, N., Haimovich, G., Barkai, O., Duek, L., and Choder, M. (2010). RNA polymerase II subunits link transcription and mRNA decay to translation. Cell *143*, 552-563.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev *12*, 357-369.

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

He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481-486.

Hekmatpanah, D.S., and Young, R.A. (1991). Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. Mol Cell Biol *11*, 5781-5791.

Hemming, S.A., Jansma, D.B., Macgregor, P.F., Goryachev, A., Friesen, J.D., and Edwards, A.M. (2000). RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. J Biol Chem *275*, 35506-35511.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci U S A *108*, 18318-18323.

Henry, N.L., Bushnell, D.A., and Kornberg, R.D. (1996). A yeast transcriptional stimulatory protein similar to human PC4. J Biol Chem *271*, 21842-21847.

Henry, N.L., Campbell, A.M., Feaver, W.J., Poon, D., Weil, P.A., and Kornberg, R.D. (1994). TFIIF-TAF-RNA polymerase II connection. Genes Dev *8*, 2868-2878.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717-728.

Hornung, G., Bar-Ziv, R., Rosin, D., Tokuriki, N., Tawfik, D.S., Oren, M., and Barkai, N. (2012). Noise-mean relationship in mutated promoters. Genome Res *22*, 2409-2417.

Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H.*, et al.* (2011). Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Res *21*, 182-192.

Houlard, M., Artus, J., Leguillier, T., Vandormael-Pournin, S., and Cohen-Tannoudji, M. (2011). DNA-RNA hybrids contribute to the replication dependent genomic instability induced by Omcg1 deficiency. Cell Cycle *10*, 108-117.

Howe, K.J., Kane, C.M., and Ares, M., Jr. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA *9*, 993-1006.

Huang, X., Wang, D., Weiss, D.R., Bushnell, D.A., Kornberg, R.D., and Levitt, M. (2010). RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. Proc Natl Acad Sci U S A *107*, 15745-15750.

Hughes, A.L., Jin, Y., Rando, O.J., and Struhl, K. (2012). A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. Mol Cell *48*, 5-15.

Hull, M.W., McKune, K., and Woychik, N.A. (1995). RNA polymerase II subunit RPB9 is required for accurate start site selection. Genes Dev *9*, 481-490.

Hyle, J.W., Shaw, R.J., and Reines, D. (2003). Functional distinctions between IMP dehydrogenase genes in providing mycophenolate resistance and guanine prototrophy to yeast. J Biol Chem *278*, 28470-28478.

Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., and Barkai, N. (2002). Revealing modular organization in the yeast transcriptional network. Nat Genet *31*, 370-377.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T., and Blencowe, B.J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. Genome Res *21*, 390-401.

Jenks, M.H., O'Rourke, T.W., and Reines, D. (2008). Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. Mol Cell Biol *28*, 3883-3893.

Jenks, M.H., and Reines, D. (2005). Dissection of the molecular basis of mycophenolate resistance in Saccharomyces cerevisiae. Yeast *22*, 1181-1190.

Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nat Struct Mol Biol *21*, 449-455.

Jin, H., and Kaplan, C.D. (2014). Relationships of RNA polymerase II genetic interactors to transcription start site usage defects and growth in Saccharomyces cerevisiae. G3 (Bethesda) *5*, 21-33.

Jin, J., Bai, L., Johnson, D.S., Fulbright, R.M., Kireeva, M.L., Kashlev, M., and Wang, M.D. (2010). Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. Nat Struct Mol Biol *17*, 745-752.

Jorgensen, P., Rupes, I., Sharom, J.R., Schneper, L., Broach, J.R., and Tyers, M. (2004). A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. Genes Dev *18*, 2491-2505.

Kahm, M., Hasenbrink, G., Lichtenberg-Frate, H., Ludwig, J., and Kschischo, M. (2010). grofit: Fitting Biological Growth Curves with R. Journal of Statistical Software *33*, 1-21.

Kamenova, I., Warfield, L., and Hahn, S. (2014). Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. Mol Cell Biol *34*, 2929-2943.

Kang, Y.N., Lai, D.P., Ooi, H.S., Shen, T.T., Kou, Y., Tian, J., Czajkowsky, D.M., Shao, Z., and Zhao, X. (2015). Genome-wide profiling of untranslated regions by paired-end ditag sequencing reveals unexpected transcriptome complexity in yeast. Mol Genet Genomics *290*, 217-224.

Kaplan, C.D. (2010). The architecture of RNA polymerase fidelity. BMC Biol *8*, 85.

Kaplan, C.D. (2013). Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae. Biochim Biophys Acta *1829*, 39-54.

Kaplan, C.D., Holland, M.J., and Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. J Biol Chem *280*, 913-922.

Kaplan, C.D., Jin, H., Zhang, I.L., and Belyanin, A. (2012). Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. PLoS Genet *8*, e1002627.

Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol Cell *30*, 547-556.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J.*, et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

Keogh, M.C., Podolny, V., and Buratowski, S. (2003). Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. Mol Cell Biol *23*, 7005-7018.

Kettenberger, H., Armache, K.J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347-357.

Kettenberger, H., Armache, K.J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol Cell *16*, 955-965.

Khaperskyy, D.A., Ammerman, M.L., Majovski, R.C., and Ponticelli, A.S. (2008). Functions of Saccharomyces cerevisiae TFIIF during transcription start site utilization. Mol Cell Biol *28*, 3757-3766.

Killeen, M., Coulombe, B., and Greenblatt, J. (1992). Recombinant TBP, transcription factor IIB, and RAP30 are sufficient for promoter recognition by mammalian RNA polymerase II. J Biol Chem *267*, 9463-9466.

Kim, H.J., Jeong, S.H., Heo, J.H., Jeong, S.J., Kim, S.T., Youn, H.D., Han, J.W., Lee, H.W., and Cho, E.J. (2004). mRNA capping enzyme activity is coupled to an early transcription elongation. Mol Cell Biol *24*, 6184-6193.

Kireeva, M.L., Nedialkov, Y.A., Cremona, G.H., Purtov, Y.A., Lubkowska, L., Malagon, F., Burton, Z.F., Strathern, J.N., and Kashlev, M. (2008). Transient reversal of RNA

polymerase II active site closing controls fidelity of transcription elongation. Molecular Cell *30*, 557-566.

Knaus, R., Pollock, R., and Guarente, L. (1996). Yeast SUB1 is a suppressor of TFIIB mutations and has homology to the human co-activator PC4. EMBO J *15*, 1933-1940.

Komissarova, N., Kireeva, M.L., Becker, J., Sidorenkov, I., and Kashlev, M. (2003). Engineering of elongation complexes of bacterial and yeast RNA polymerases. Methods Enzymol *371*, 233-251.

Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature *462*, 323-330.

Koyama, H., Ito, T., Nakanishi, T., and Sekimizu, K. (2007). Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. Genes Cells *12*, 547-559.

Kuehner, J.N., and Brow, D.A. (2006). Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. Journal of Biological Chemistry *281*, 14119-14128.

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.

Kugel, J.F., and Goodrich, J.A. (2002). Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. Mol Cell Biol *22*, 762-773.

Kulaeva, O.I., Gaykalova, D.A., Pestov, N.A., Golovastov, V.V., Vassylyev, D.G., Artsimovitch, I., and Studitsky, V.M. (2009). Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. Nat Struct Mol Biol *16*, 1272-1278.

Kulaeva, O.I., Hsieh, F.K., and Studitsky, V.M. (2010). RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. Proc Natl Acad Sci U S A *107*, 11325-11330.

Kulish, D., and Struhl, K. (2001). TFIIS enhances transcriptional elongation through an artificial arrest site in vivo. Mol Cell Biol *21*, 4162-4168.

Kwapisz, M., Wery, M., Despres, D., Ghavi-Helm, Y., Soutourina, J., Thuriaux, P., and Lacroute, F. (2008). Mutations of RNA polymerase II activate key genes of the nucleoside triphosphate biosynthetic pathways. EMBO J *27*, 2411-2421.

Lada, A.G., Stepchenkova, E.I., Waisertreiger, I.S., Noskov, V.N., Dhar, A., Eudy, J.D., Boissy, R.J., Hirano, M., Rogozin, I.B., and Pavlov, Y.I. (2013). Genome-wide mutation

avalanches induced in diploid yeast cells by a base analog or an APOBEC deaminase. PLoS Genet *9*, e1003736.

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

Larson, M.H., Zhou, J., Kaplan, C.D., Palangat, M., Kornberg, R.D., Landick, R., and Block, S.M. (2012). Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. Proc Natl Acad Sci U S A *109*, 6555-6560.

Laufer, C., Fischer, B., Billmann, M., Huber, W., and Boutros, M. (2013). Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. Nat Methods *10*, 427-431.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.

Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A., Margaritis, T., van de Pasch, L.A., van Heesch, S.A., Brok, M.O., Groot Koerkamp, M.J.*, et al.* (2011). The specificity and topology of chromatin interaction pathways in yeast. Mol Cell *42*, 536-549.

Li, H., Hou, J., Bai, L., Hu, C., Tong, P., Kang, Y., Zhao, X., and Shao, Z. (2015). Genome-wide Analysis of Core Promoter Structures in Schizosaccharomyces pombe with DeepCAGE. RNA Biol, 0.

Li, X.Y., Bhaumik, S.R., and Green, M.R. (2000). Distinct classes of yeast promoters revealed by differential TAF recruitment. Science *288*, 1242-1244.

Li, Y., Flanagan, P.M., Tschochner, H., and Kornberg, R.D. (1994). RNA polymerase II initiation factor interactions and transcription start site selection. Science *263*, 805-807.

Lin, Y.Y., Kiihl, S., Suhail, Y., Liu, S.Y., Chou, Y.H., Kuang, Z., Lu, J.Y., Khor, C.N., Lin, C.L., Bader, J.S.*, et al.* (2012). Functional dissection of lysine deacetylases reveals that HDAC1 and p300 regulate AMPK. Nature *482*, 251-255.

Lindstrom, D.L., and Hartzog, G.A. (2001). Genetic interactions of Spt4-Spt5 and TFIIS with the RNA polymerase II CTD and CTD modifying enzymes in Saccharomyces cerevisiae. Genetics *159*, 487-497.

Lipson, D., Raz, T., Kieu, A., Jones, D.R., Giladi, E., Thayer, E., Thompson, J.F., Letovsky, S., Milos, P., and Causey, M. (2009). Quantification of the yeast transcriptome by single-molecule sequencing. Nat Biotechnol *27*, 652-658.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J Biomed Biotechnol *2012*, 251364.

Liu, X., Bushnell, D.A., Silva, D.A., Huang, X., and Kornberg, R.D. (2011). Initiation complex structure and promoter proofreading. Science *333*, 633-637.

Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. Mol Cell Biol *24*, 1721-1735.

Lubliner, S., Keren, L., and Segal, E. (2013). Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res *41*, 5569-5581.

Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., and Segal, E. (2015). Core promoter sequence in yeast is a major determinant of expression level. Genome Res.

Lue, N.F., Flanagan, P.M., Sugimoto, K., and Kornberg, R.D. (1989). Initiation by Yeast Rna Polymerase-Ii at the Adenoviral Major Late Promoter Invitro. Science *246*, 661-664.

Luse, D.S. (2012). Rethinking the role of TFIIF in transcript initiation by RNA polymerase II. Transcription *3*, 156-159.

Majovski, R.C., Khaperskyy, D.A., Ghazy, M.A., and Ponticelli, A.S. (2005). A functional role for the switch 2 region of yeast RNA polymerase II in transcription start site utilization and abortive initiation. J Biol Chem *280*, 34917-34923.

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by Nonsense-Mediated-mRNA Decay. Elife *4*.

Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., and Strathern, J.N. (2006). Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. Genetics *172*, 2201-2209.

Malinen, A.M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M.S., and Belogurov, G.A. (2012). Active site opening and closure control translocation of multisubunit RNA polymerase. Nucleic Acids Res *40*, 7442-7451.

Malone, E.A., Clark, C.D., Chiang, A., and Winston, F. (1991). Mutations in SPT16/CDC68 suppress cis- and trans-acting mutations that affect promoter function in Saccharomyces cerevisiae. Mol Cell Biol *11*, 5710-5717.

Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol Cell *17*, 831-840.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res *18*, 1073-1083.

McManus, C.J., May, G.E., Spealman, P., and Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res *24*, 422-430.

Miller, G., and Hahn, S. (2006). A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. Nat Struct Mol Biol *13*, 603-610.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J.*, et al.* (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. Science *342*, 1238724.

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 *320*, 1344-1349.

Nayak, D., Voss, M., Windgassen, T., Mooney, R.A., and Landick, R. (2013). Cys-pair reporters detect a constrained trigger loop in a paused RNA polymerase. Mol Cell *50*, 882-893.

Nelson, H.C., Finch, J.T., Luisi, B.F., and Klug, A. (1987). The structure of an oligo(dA).oligo(dT) tract and its biological implications. Nature *330*, 221-226.

Nesser, N.K., Peterson, D.O., and Hawley, D.K. (2006). RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. Proceedings of the National Academy of Sciences of the United States of America *103*, 3268-3273.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature *377*, 119-128.

Niyogi, S.K., and Feldman, R.P. (1981). Effect of several metal ions on misincorporation during transcription. Nucleic Acids Res *9*, 2615-2627.

Nonet, M., Sweetser, D., and Young, R.A. (1987). Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell *50*, 909-915.

Ohkuni, K., and Kitagawa, K. (2011). Endogenous transcription at the centromere facilitates centromere activity in budding yeast. Curr Biol *21*, 1695-1703.

Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M., and Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature *400*, 284-288.

Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (1999). Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. Molecular cell *3*, 109-118.

Pal, M., Ponticelli, A.S., and Luse, D.S. (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. Mol Cell *19*, 101-110.

Pan, G., and Greenblatt, J. (1994). Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. J Biol Chem *269*, 30101-30104.

Pan, X., Yuan, D.S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J.S., Hieter, P., Spencer, F., and Boeke, J.D. (2004). A robust toolkit for functional profiling of the yeast genome. Mol Cell *16*, 487-496.

Pappas, D.L., Jr., and Hampsey, M. (2000). Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in Saccharomyces cerevisiae. Mol Cell Biol *20*, 8343-8351.

Park, D., Morris, A.R., Battenhouse, A., and Iyer, V.R. (2014). Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. Nucleic Acids Res *42*, 3736-3749.

Parvin, J.D., and Sharp, P.A. (1993). DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. Cell *73*, 533-540.

Pascual-Garcia, P., Govind, C.K., Queralt, E., Cuenca-Bono, B., Llopis, A., Chavez, S., Hinnebusch, A.G., and Rodriguez-Navarro, S. (2008). Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. Genes Dev *22*, 2811-2822.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127-131.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell *36*, 178-191.

Pikielny, C.W., and Rosbash, M. (1985). mRNA splicing efficiency in yeast and the contribution of nonconserved sequences. Cell *41*, 119-126.

Pinto, I., Ware, D.E., and Hampsey, M. (1992). The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo. Cell *68*, 977-988.

Pinto, I., Wu, W.H., Na, J.G., and Hampsey, M. (1994). Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. J Biol Chem *269*, 30569-30573.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. (2007). Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. PLoS Biol *5*, e90.

Plessy, C., Bertin, N., Takahashi, H., Simone, R., Salimullah, M., Lassmann, T., Vitezic, M., Severin, J., Olivarius, S., Lazarevic, D*., et al.* (2010). Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. Nat Methods *7*, 528-534.

Powell, W., and Reines, D. (1996). Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro. J Biol Chem *271*, 6866-6873.

Prelich, G., and Winston, F. (1993). Mutations that suppress the deletion of an upstream activating sequence in yeast: involvement of a protein kinase and histone H3 in repressing transcription in vivo. Genetics *135*, 665-676.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Ranish, J.A., and Hahn, S. (1991). The yeast general transcription factor TFIIA is composed of two polypeptide subunits. J Biol Chem *266*, 19320-19327.

Rao, Y., Lee, Y., Jarjoura, D., Ruppert, A.S., Liu, C.G., Hsu, J.C., and Hagan, J.P. (2008). A comparison of normalization techniques for microRNA microarray data. Stat Appl Genet Mol Biol *7*, Article22.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol *14*, R95.

Reines, D. (2003). Use of RNA yeast polymerase II mutants in studying transcription elongation. Methods Enzymol *371*, 284-292.

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

Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.

Riles, L., Shaw, R.J., Johnston, M., and Reines, D. (2004). Large-scale screening of yeast mutants for sensitivity to the IMP dehydrogenase inhibitor 6-azauracil. Yeast *21*, 241-248.

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

Roguev, A., Talbot, D., Negri, G.L., Shales, M., Cagney, G., Bandyopadhyay, S., Panning, B., and Krogan, N.J. (2013). Quantitative genetic-interaction mapping in mammalian cells. Nat Methods *10*, 432-437.

Roguev, A., Wiren, M., Weissman, J.S., and Krogan, N.J. (2007). High-throughput genetic interaction mapping in the fission yeast Schizosaccharomyces pombe. Nat Methods *4*, 861-866.

Rojas-Duran, M.F., and Gilbert, W.V. (2012). Alternative transcription start site selection leads to large differences in translation activity in yeast. RNA *18*, 2299-2305.

Rosin, D., Hornung, G., Tirosh, I., Gispan, A., and Barkai, N. (2012). Promoter nucleosome organization shapes the evolution of gene expression. PLoS Genet *8*, e1002579.

Rosonina, E., Willis, I.M., and Manley, J.L. (2009). Sub1 functions in osmoregulation and in transcription by both RNA polymerases II and III. Mol Cell Biol *29*, 2308-2321.

Russell, P.R. (1983). Evolutionary divergence of the mRNA transcription initiation mechanism in yeast. Nature *301*, 167-169.

Ryan, C.J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S.R.*, et al.* (2012). Hierarchical Modularity and the Evolution of Genetic Interactomes across Species. Mol Cell *46*, 691-704.

Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature *493*, 437-440.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol *7*, 557-567.

Schmitt, M.E., Brown, T.A., and Trumpower, B.L. (1990). A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic Acids Res *18*, 3091-3092.

Schneider, S., Pei, Y., Shuman, S., and Schwer, B. (2010). Separable functions of the fission yeast Spt5 carboxyl-terminal domain (CTD) in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. Molecular and cellular biology *30*, 2353-2364.

Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F.*, et al.* (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell *123*, 507-519.

Schuldiner, M., Collins, S.R., Weissman, J.S., and Krogan, N.J. (2006). Quantitative genetic analysis in Saccharomyces cerevisiae using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. Methods *40*, 344-352.

Schwabish, M.A., and Struhl, K. (2007). The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. Mol Cell Biol *27*, 6987-6995.

Schweikhard, V., Meng, C., Murakami, K., Kaplan, C.D., Kornberg, R.D., and Block, S.M. (2014). Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. Proc Natl Acad Sci U S A *111*, 6642-6647.

Seizl, M., Hartmann, H., Hoeg, F., Kurth, F., Martin, D.E., Soding, J., and Cramer, P. (2011). A conserved GA element in TATA-less RNA polymerase II promoters. PLoS One *6*, e27595.

Shaw, R.J., and Reines, D. (2000). Saccharomyces cerevisiae transcription elongation mutants are defective in PUR5 induction in response to nucleotide depletion. Mol Cell Biol *20*, 7427-7437.

Shaw, R.J., Wilson, J.L., Smith, K.T., and Reines, D. (2001). Regulation of an IMP dehydrogenase gene and its overexpression in drug-sensitive transcription elongation mutants of yeast. J Biol Chem *276*, 32905-32916.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Sikorski, T.W., Ficarro, S.B., Holik, J., Kim, T., Rando, O.J., Marto, J.A., and Buratowski, S. (2011). Sub1 and RPA associate with RNA polymerase II at different stages of transcription. Mol Cell *44*, 397-409.

Silva, D.A., Weiss, D.R., Pardo Avila, F., Da, L.T., Levitt, M., Wang, D., and Huang, X. (2014). Millisecond dynamics of RNA polymerase II translocation at atomic resolution. Proc Natl Acad Sci U S A *111*, 7665-7670.

Simchen, G., Winston, F., Styles, C.A., and Fink, G.R. (1984). Ty-mediated gene expression of the LYS2 and HIS4 genes of Saccharomyces cerevisiae is controlled by the same SPT genes. Proc Natl Acad Sci U S A *81*, 2431-2434.

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

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.

Song, Y.H., and Ahn, S.H. (2010). A Bre1-associated protein, large 1 (Lge1), promotes H2B ubiquitylation during the early stages of transcription elongation. J Biol Chem *285*, 2361-2367.

Spahr, H., Calero, G., Bushnell, D.A., and Kornberg, R.D. (2009). Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. Proceedings of the National Academy of Sciences of the United States of America *106*, 9185-9190.

Spencer, F., Gerring, S.L., Connelly, C., and Hieter, P. (1990). Mitotic chromosome transmission fidelity mutants in Saccharomyces cerevisiae. Genetics *124*, 237-249.

Stein, H., and Hausen, P. (1969). Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. Science *166*, 393-395.

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.

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A *90*, 6498-6502.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev *26*, 163-175.

Struhl, K. (1987). Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell *49*, 295-297.

Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. Annual review of biochemistry *58*, 1051-1077.

Sugihara, F., Kasahara, K., and Kokubo, T. (2011). Highly redundant function of multiple AT-rich sequences as core promoter elements in the TATA-less RPS5 promoter of Saccharomyces cerevisiae. Nucleic Acids Res *39*, 59-75.

Sun, M., Lariviere, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). The Journal of biological chemistry *285*, 41597-41603.

Sun, Z.W., and Hampsey, M. (1995). Identification of the gene (SSU71/TFG1) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *92*, 3127-3131.

Sun, Z.W., and Hampsey, M. (1996). Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. Mol Cell Biol *16*, 1557-1566.

Sun, Z.W., Tessmer, A., and Hampsey, M. (1996). Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in Saccharomyces cerevisiae. Nucleic Acids Res *24*, 2560-2566.

Svejstrup, J.Q., Wang, Z., Feaver, W.J., Wu, X., Bushnell, D.A., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1995). Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. Cell *80*, 21-28.

Svetlov, V., and Nudler, E. (2013). Basic mechanism of transcription by RNA polymerase II. Biochim Biophys Acta *1829*, 20-28.

Svetlov, V., Vassylyev, D.G., and Artsimovitch, I. (2004). Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. J Biol Chem *279*, 38087-38090.

Sweeney, M.J. (1977). Mycophenolic acid and its mechanism of action in cancer and psoriasis. Jpn J Antibiot *30 Suppl*, 85-92.

Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol Cell *34*, 710-721.

Tagami, S., Sekine, S., Kumarevel, T., Hino, N., Murayama, Y., Kamegamori, S., Yamamoto, M., Sakamoto, K., and Yokoyama, S. (2010). Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. Nature *468*, 978-982.

Tan, L., Wiesler, S., Trzaska, D., Carney, H.C., and Weinzierl, R.O. (2008). Bridge helix and trigger loop perturbations generate superactive RNA polymerases. J Biol *7*, 40.

Tan, Q., Prysak, M.H., and Woychik, N.A. (2003). Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. Mol Cell Biol *23*, 3329-3338.

Tavenet, A., Suleau, A., Dubreuil, G., Ferrari, R., Ducrot, C., Michaut, M., Aude, J.C., Dieci, G., Lefebvre, O., Conesa, C.*, et al.* (2009). Genome-wide location analysis reveals a role for Sub1 in RNA polymerase III transcription. Proc Natl Acad Sci U S A *106*, 14265-14270.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife *2*, e00534.

Temiakov, D., Zenkin, N., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., Kashkina, E., Savkina, M., Zorov, S., Nikiforov, V., Igarashi, N.*, et al.* (2005). Structural basis of transcription inhibition by antibiotic streptolydigin. Mol Cell *19*, 655-666.

Thiebaut, M., Colin, J., Neil, H., Jacquier, A., Seraphin, B., Lacroute, F., and Libri, D. (2008). Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae. Molecular cell *31*, 671-682.

Timmers, H.T. (1994). Transcription initiation by RNA polymerase II does not require hydrolysis of the beta-gamma phosphoanhydride bond of ATP. EMBO J *13*, 391-399.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res *18*, 1084-1091.

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M.*, et al.* (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808-813.

Toulokhonov, I., Zhang, J., Palangat, M., and Landick, R. (2007). A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. Mol Cell *27*, 406-419.

Trcek, T., Larson, D.R., Moldon, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell *147*, 1484-1497.

Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L.*, et al.* (2008). High-throughput, quantitative analyses of genetic interactions in E. coli. Nat Methods *5*, 781-787.

Tyree, C.M., George, C.P., Lira-DeVito, L.M., Wampler, S.L., Dahmus, M.E., Zawel, L., and Kadonaga, J.T. (1993). Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. Genes Dev *7*, 1254-1265.

van Bakel, H., Tsui, K., Gebbia, M., Mnaimneh, S., Hughes, T.R., and Nislow, C. (2013). A compendium of nucleosome and transcript profiles reveals determinants of chromatin architecture and transcription. PLoS Genet *9*, e1003479.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D., and Holstege, F.C. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. EMBO Rep *4*, 387-393.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends Genet *30*, 418-426.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. Nature *448*, 163-168.

Viktorovskaya, O.V., Engel, K.L., French, S.L., Cui, P., Vandeventer, P.J., Pavlovic, E.M., Beyer, A.L., Kaplan, C.D., and Schneider, D.A. (2013). Divergent contributions of conserved active site residues to transcription by eukaryotic RNA polymerases I and II. Cell Rep *4*, 974-984.

Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol Cell *44*, 978-988.

Walmacq, C., Kireeva, M.L., Irvin, J., Nedialkov, Y., Lubkowska, L., Malagon, F., Strathern, J.N., and Kashlev, M. (2009). Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. The Journal of biological chemistry *284*, 19601-19612.

Wang, B., Predeus, A.V., Burton, Z.F., and Feig, M. (2013). Energetic and structural details of the trigger-loop closing transition in RNA polymerase II. Biophys J *105*, 767-775.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell *127*, 941-954.

Wang, W., Carey, M., and Gralla, J.D. (1992). Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. Science *255*, 450-453.

Weilbaecher, R., Hebron, C., Feng, G., and Landick, R. (1994). Termination-altering amino acid substitutions in the beta' subunit of Escherichia coli RNA polymerase identify regions involved in RNA chain elongation. Genes Dev *8*, 2913-2927.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res *20*, 90-100.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481-489.

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.

Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics *107*, 179-197.

Winston, F., and Sudarsanam, P. (1998). The SAGA of Spt proteins and transcriptional analysis in yeast: past, present, and future. Cold Spring Harb Symp Quant Biol *63*, 553-561.

Wittschieben, B.O., Otero, G., de Bizemont, T., Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C.D., Tempst, P., and Svejstrup, J.Q. (1999). A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. Molecular cell *4*, 123-128.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol Cell *54*, 601-612.

Wu, J., Awrey, D.E., Edwards, A.M., Archambault, J., and Friesen, J.D. (1996). In vitro characterization of mutant yeast RNA polymerase II with reduced binding for elongation factor TFIIS. Proc Natl Acad Sci U S A *93*, 11552-11557.

Wu, R., and Li, H. (2010). Positioned and G/C-capped poly(dA:dT) tracts associate with the centers of nucleosome-free regions in yeast promoters. Genome Res *20*, 473-484.

Wu, W.H., Pinto, I., Chen, B.S., and Hampsey, M. (1999). Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB. Genetics *153*, 643-652.

Xi, Y., Yao, J., Chen, R., Li, W., and He, X. (2011). Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. Genome Res *21*, 718-724.

Xiao, T., Kao, C.F., Krogan, N.J., Sun, Z.W., Greenblatt, J.F., Osley, M.A., and Strahl, B.D. (2005). Histone H2B ubiquitylation is associated with elongating RNA polymerase II. Mol Cell Biol *25*, 637-651.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442-451.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626-630.

Yuen, K.W., Warren, C.D., Chen, O., Kwok, T., Hieter, P., and Spencer, F.A. (2007). Systematic genome instability screens in yeast and their potential relevance to cancer. Proc Natl Acad Sci U S A *104*, 3925-3930.

Yuzenkova, Y., Bochkareva, A., Tadigotla, V.R., Roghanian, M., Zorov, S., Severinov, K., and Zenkin, N. (2010). Stepwise mechanism for transcription fidelity. BMC Biol *8*, 54.

Yuzenkova, Y., and Zenkin, N. (2010). Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. Proc Natl Acad Sci U S A *107*, 10878-10883.

Zentner, G.E., and Henikoff, S. (2013). Mot1 redistributes TBP from TATA-containing to TATA-less promoters. Mol Cell Biol *33*, 4996-5004.

Zhang, D.Y., Carson, D.J., and Ma, J. (2002). The role of TFIIB-RNA polymerase II interaction in start site selection in yeast cells. Nucleic Acids Res *30*, 3078-3085.

Zhang, J., Palangat, M., and Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. Nat Struct Mol Biol *17*, 99-104.

Zhang, Y., Degen, D., Ho, M.X., Sineva, E., Ebright, K.Y., Ebright, Y.W., Mekler, V., Vahedian-Movahed, H., Feng, Y., Yin, R.*, et al.* (2014). GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. Elife *3*, e02450.

Zhang, Y., Shishkin, A.A., Nishida, Y., Marcinkowski-Desmond, D., Saini, N., Volkov, K.V., Mirkin, S.M., and Lobachev, K.S. (2012). Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. Mol Cell *48*, 254-265.

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. Nucleic Acids Res *33*, 2838-2851.

APPENDIX A

OPPOSITE METAL SENSITIVITY CONFERRED BY POL II MUTANTS

I showed that Pol II activity mutants have allele specific sensitivities and resistance to divalent cations specifically Manganese. Manganese can function in Pol II active sites as replacement of Magnesium, but may alter Pol II activity. Different metals may also alter fidelity of transcription process. I performed piorneer phenotyping analysis for our Pol II activity mutants utilizing our genetic interaction assays. Observations of how Manganese (Mn) affect growth defects in Pol II allele specific manner are contributed to publication of (Cabart et al., 2014). Other data including suppression of Mn by Magnesium (Mg), Cobalt (Co) effect on Pol II mutants and Mg effect on those, Mn effect on TSS at *ADH1* etc. were contributed to developing a thesis project of a graduate student in the lab.

Mn and Mg are essential cations for many enzymatic reactions. Previous studies showed that Mn can replace Mg in Pol II active site as the essential metal during transcription elongation and suggested Mn could alter transcription in several facets (Niyogi and Feldman, 1981; Walmacq et al., 2009).

This project started with an interesting observation made by Kaplan lab in University of Utah with our Pol II activity mutant strains. In a preliminary spot growth assay with our Pol II strains, Liangtao Li in Kaplan lab observed sensitivity of two strong fast mutants E1103G and G1097D on medium containing 15mM Mn and no apparent phenotype for two slow mutants. The Mn sensitivities of two fast mutants were

both suppressed by a strong slow mutant H1085Y. I expanded the experiments with a set of Pol II activity mutants with series of concentration with Mn, and I show the results in **Figure A-1**.

Fast mutants showed increasing sensitivity in the presence of series of concentrations of Mn, and the sensitivity correlates their magnitude of activity alterations. The strongest fast mutant G1097D showed lethality starting at a relatively low concentration of Mn (5~10mM). Another strong fast mutant E1103G is lethal at a concentration higher than 15mM Mn (for example, 18mM Mn). Conversely, slow mutants show opposite phenotypes- Mn suppresses growth defects of theirs (compare mutant and WT growth differences in plain media and with 15mM Mn) (**Figure A-1**). We have seen a rescue of lethality in Pol II loss of function mutant (LOF)s previously in combination of fast and slow lethal mutants (Kaplan et al., 2012), thus the hypothesis for Mn was that it altered Pol II activity. I tested a few lethal LOF mutants, and observed two lethal LOFs could be rescued by high concentration of Mn (18~20mM Mn), however their growth was severely defective for downstream phenotyping. Addition of Mg in the media at the presence of Mn suppressed growth defect of weak fast mutants but could not rescue lethality of strongest fast mutant G1097D.

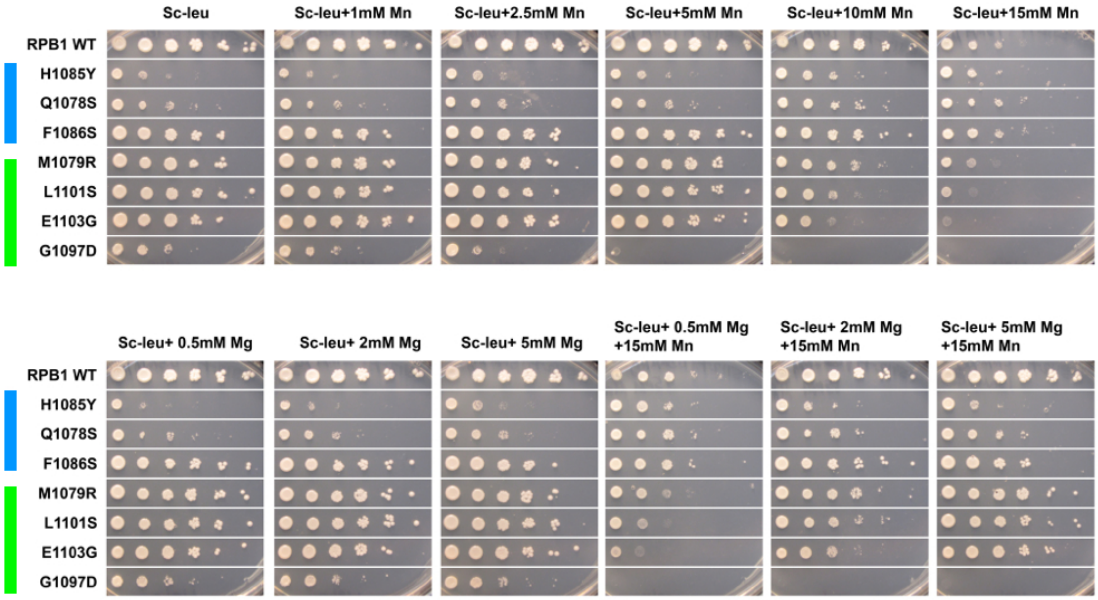


Figure A-1. Serial dilutions of Pol II *rpo21/rpb1* alleles on selective medium and medium with Manganese and/or Magnesium.

Fast mutants are marked by the green bar, slow mutants are marked by the blue bar. Final concentrations of metal ions in medium are shown.

In addition, Mg removed the suppression of slow mutants' growth defect by Mn. If Mn affected Pol II catalytic activity, we would be able to detect its effect on *ADH1* TSS utilization as all of our previous results showed tight correlation between altered catalytic activities and TSS utilization at model gene *ADH1* (Chapter II, III, IV). Therefore, I performed the primer extension assay at *ADH1* in WT and a strong fast mutant E1103G grown in media (same SC media as in spot growth assay) containing 10-15mM Mn. Indeed, Mn conferred moderate upstream shift in WT, the magnitude of shifts did not vary in the 10-15mM concentration of Mn. It also exaggerated upstream shift of E1103G moderately, *i.e.,* Mn effect on E1103G mutant was similar extent as its effect on WT.

Mn was as shown to decrease transcription fidelity in *in vitro* transcription (Niyogi and Feldman, 1981; Walmacq et al., 2009). Transcription infidelity presumably has broad impact on multi-level cellular functions. Pol II fast mutant E1103G has been shown to decrease transcription fidelity (Kireeva et al., 2008; Larson et al., 2012) and has various growth phenotypes *in vivo*, consistent with a model transcription infidelity can cause growth defect *in vivo*. We asked if *dst1*Δ (deletion of TFIIS, elongation factor that promotes transcription proof reading by cleavage of misincorporated nucleotide) conferred Mn sensitivity. If decreased fidelity caused by Mn is exacerbated in the fast mutants, and is a main contributor of the exaggerated growth defects of the fast mutants, one expects *dst1*Δ to show similar Mn sensitivity in growth. We have shown that *dst1*Δ shows synthetic lethality with strong fast mutants (Chapter II, III), and our current model suggests this is through increased transcription errors deteriorating cellular fitness.

However, we did not see any apparent changes in *dst1*Δ growth when Mn was added in the media. This contrasts the simple model that Mn causes exacerbated growth defects of Pol II fast mutants by increasing their transcription infidelity.

Currently it is not clear what is the main cause of Mn phenotypes in Pol II mutants, Chenxi Qiu in Kaplan lab will continue the investigation as one of his thesis projects.

APPENDIX B

INFLUENCE OF POL II ACTIVITY ON R LOOP FORMATION IN THE TRANSCRIPTION PROCESS

To test how Pol II mutants relate to other transcription defects and the effects on cell growth and other conditional growth phenotypes, I studied how Pol II activity mutants influence other aspects of transcription including R loop formation in the transcription process.

R loops are nucleic acid structure formed by RNA:DNA hybrid and a displaced DNA strand. They possibly arise from transcripts threading back before DNAs can anneal back to from the duplex (thread back model), and are reported to occur preferentially in G-rich region (Yu et al., 2003). The resulting three stranded nucleic acid structure R loops pose a potential threat to genome stability. R loop possibly causes replication and transcription machinery collision, posing barriers to DNA replication and subsequently promoter recombination due to the unexpected exposed ssDNA (Aguilera and Garcia-Muse, 2012; Houlard et al., 2011). The exposed ssDNA is also a vulnerable subject for DNA damaging agents and for repair mechanisms leading to lesions and nicks. However, more recently R loops have also been suggested to function as a regulator of genome integrity and gene expression by recruiting histone methyltransferases or DNA demethylases (Ginno et al., 2012). It has also been proposed that R loop promotes transcription termination when the level of R loop is balanced well in human (Skourti-Stathaki and Proudfoot, 2014; Skourti-Stathaki et al., 2011). A critical

surveillance system for R loops is digestion of RNA in RNA-DNA hybrid by enzyme RNase H (Cerritelli and Crouch, 2009; Stein and Hausen, 1969). Deletions of RNase H alleles have been shown to result in increased genome instability in *S. cerevisiae* (Stirling et al., 2012; Wahba et al., 2011). Overexpression of RNase H encoding alleles has been used to remove cellular level of R loops in previous studies (Cerritelli et al., 2003).

To examine how R loop affects our Pol II phenotypes, I combined Pol II mutants with *RNH1* and *RNH201* (genes encoding alleles of RNase H) constructed in high copy vector pRS423 to examine genetic interactions between RNH alleles and Pol II mutants with our growth phenotypes. If R loop contributes to phenotypes we observe in our Pol II mutants, by overexpression of one or both RNase H coding alleles, we expected to see the alleviation of the phenotypes. I had three different sets of constructs including overexpression (pRS423) of RNH1, RNH201 and both RNH1 and RNH201 combined with Pol II WT and mutants. However, I did not observe any notable phenotypes in my patch growth assay and spots assay. A graduate student Indranil Malik in the lab discovered later that the same high copy vector did not "over" express his target in very similar setup as in my experiments. He showed that, in contrast to 20 fold difference one expects, high copy vector merely gave rise to ~2−3 fold expression compared to control by western blot. To perform the experiments in a correct way, a true high copy vector that can overexpress RNase H alleles are needed.