Determination of DNA sequence, Pol II catalytic activity and TFIIH subunit Tfb3 functions in transcription start site selection in *Saccharomyces cerevisiae*

by

Yunye Zhu

B.S., Beijing Forestry University, Beijing, China, 2012

M.S., Hefei University of Technology, Hefei, Anhui, China, 2015

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yunye Zhu

It was defended on

December 7, 2021

and approved by

Sarah Hainer, PhD, Assistant Professor, Department of Biological Sciences

Jeffrey L. Brodsky, PhD, Professor, Department of Biological Sciences

Mark Rebeiz, PhD, Associate Professor, Department of Biological Sciences

Anne-Ruxandra Carvunis, PhD, Assistant Professor, Department of Computational and Systems Biology

Dissertation Director: Craig D. Kaplan, PhD, Associate Professor, Department of Biological Sciences

Determination of DNA sequence, Pol II catalytic activity and TFIIH subunit Tfb3 functions in transcription start site selection in *Saccharomyces cerevisiae*

Yunye Zhu, PhD

University of Pittsburgh, 2022

During initiation of transcription by RNA Polymerase II (Pol II) where and how efficiently transcription initiates is determined by a constellation of inputs including DNA sequence and activities of Pol II and General Transcription Factors (GTFs). Not all promoter positions allow initiation as only subsets of DNA sequences are able to be used as Transcription Start Sites (TSSs). Furthermore, the "promoter scanning" process that determines TSS usage in *Saccharomyces cerevisiae* is sensitive to alterations in Pol II and GTF factor activity. My dissertation focuses on three determinants of TSS selection: DNA sequence, Pol II catalytic activity and GTF TFIIH subunit Tfb3. To dissect Pol II TSS sequence specificity and how Pol II activity alters it, I develop a massively parallel reporter assay "Pol II MASTER" (Pol II MAssively Systematic Transcript End Readout). Using Pol II MASTER, I measure the efficiency of transcription initiation during promoter scanning by *S. cerevisiae* Pol II for ~1 million unique TSS sequences. Pol II MASTER not only recapitulates known critical qualities of *S. cerevisiae* TSS -8, -1, and +1 positions but also demonstrates that surrounding sequences modulate initiation efficiency over a wide range. I discover functional interactions between neighboring sequence positions, indicating that adjacent positions likely function together. These results enable development of a predictable model for initiation efficiency at genomic promoters. I demonstrate that Pol II mutants with altered catalytic activity selectively modulate preference for initiating nucleotide. To determine how Tfb3 functions in promoter scanning and interacts with other initiation factors. I identified two classes of *tfb3*

alleles that confer polar effects on TSS usage, namely they can shift TSS usage either upstream or downstream. I find that *tfb3* allele effects on TSS usage show an intermediate pattern between Pol II efficiency alleles and TFIIH processivity alleles, suggesting Tfb3 might function in TSS selection through both Pol II efficiency and TFIIH processivity. I show primary additive genetic interactions between *tfb3* alleles and alleles of TFIIH DNA translocase subunit Ssl2 or initiation cofactor Sub1, suggesting Tfb3 functions in regulating TFIIH's processivity.

Table of contents

List of tables

List of figures

Preface

The past six and half years have been a long journey for me to get here. The completion of my dissertation would not have been possible without the support from many people.

I would like to express my deepest appreciation to my advisor Dr. Craig D. Kaplan for his invaluable guidance, enduring support, and encouragement through each stage of the process. I am always impressed by his passion for science, thorough knowledge of many aspects, high standards on research, and willingness to guide students. I am thankful for his feedback and patience for motivating me to recognize what I have done well and what I can improve to get to the next level. He spent a lot of time helping me with critical thinking, presentation/writing skills, and project management. I am so grateful for his encouragement and mentorship helping me to build confidence and shaping me into a better person. I still remember how happy I was when I saw my name was included in the "Kaplan Transcription Factory" lab door sign, the moment I knew I was accepted into the Kaplan lab. Joining the Kaplan lab is one of the best decisions in my life.

I would like to extend my deepest gratitude to my committee members Dr. Sarah Hainer, Dr. Jeffrey L. Brodsky, Dr. Mark Rebeiz, Dr. Anne-Ruxandra Carvunis at University of Pittsburgh and Dr. Michael Polymenis, Dr. Lanying Zeng, Dr. Jerome Menet at Texas A&M University for sharing their invaluable insights and providing constructive discussions and feedback on my projects. I would especially like to thank Dr. Sarah Hainer for her great support and encouragement. I am also grateful to Dr. Anne-Ruxandra Carvunis for her valuable suggestions on data analysis.

I wish to thank many faculty and staff for providing guidance, sharing valuable resources and assistance. I would like to thank Dr. Karen Arndt, Dr. Miler Lee, and Dr. Mary Bryk, Dr.

David Peterson (Texas A&M University) and their lab members for insightful discussions and feedback during supergroup meetings and for sharing valuable resources. I would like to thank Dr. Hays Rye, Dr. Jennifer Herman, and Dr. Alan R. Dabney (Texas A&M University) for their exceptional teaching in biology or statistics courses. I would like to thank Dr. Jerome Menet and Dr. Anthony Schwacha for sharing their protocols, equipment, and reagents. I would like to thank Dr. Margaret Glasner, Dr. John Mullet, and Dr. Dorothy Shippen (Texas A&M University) for their mentoring during my rotations. I would like to thank Shuoran Li for discussions with the Statistical Consulting Center. I would like to appreciate the assistance that I received from Cathy Barr and Rafael R. Almanzar (Texas A&M University).

I would like to thank our collaborators Dr. Bryce E. Nickels and Dr. Irina O. Vvedenskaya (Rutgers University) for their work on TSS-seq and their contribution to manuscript preparation.

I'm deeply indebted to Kaplan lab former and current members. I would like to thank my former colleagues Dr. Huiyan Jin and Dr. Indranil Malik for their early guidance and support. I am extremely grateful to Dr. Chenxi Qiu for his invaluable advice and "CQ-style" support from my first day in the Kaplan lab until now. I would also like to extend my gratitude to Dr. Tingting Zhao for her extensive guidance and support in both science and personal life. I would like to thank Dr. Anandhakumar Jayamani and current colleagues Bingbing Duan and Payal Arora for their all the time support from College Station to Pittsburgh, especially during the pandemic. I also wish to thank my other colleagues Evelyn Callaway, Shrabani Basu, Pratik Basnet, and Alex Visbisky for their help.

Last but not least, my deepest gratitude goes to my parents. Without their unwavering love and unfailing support, I would not be in the position I am today. In addition, I could not have

completed this journey without support of my friends. Many thanks to my friends for being there for me and for the strength they gave me.

1.0 Introduction

DNA-dependent RNA Polymerases (Pols) synthesize RNA from a DNA template through the process of transcription. Eukaryotic nuclear RNA Pols (including Pol I, Pol II, and Pol III in all eukaryotes, and additional Pol IV and Pol V in plants) are highly conserved, and this conservation extends to components (General Transcription Factors, GTFs) that are required to for initiating transcription. For Pol I and III, some of these factors have evolved to be incorporated into the enzyme complexes themselves, while Pol II, the focus of this dissertation, has GTFs that assemble with Pol II specifically for initiation. Furthermore, Pol II is unique among eukaryotic nuclear RNA Pols in requiring function of an ATPase for promoter opening and Transcription Start Site (TSS) selection (Bunick et al., 1982; Holstege et al., 1997; Wang et al., 1992). These functional differences may relate to the much more extensive and diverse regulation of Pol II initiation relative to the other eukaryotic Pols.

Pol II transcribes all protein-coding genes and many small nuclear RNAs in three sequential transcription steps: initiation, elongation, and termination. The first step in transcription, initiation, determines where and how efficiently transcription initiates and therefore is a critical point of control in gene expression. Pol II alone cannot bind to promoter DNA and precisely initiate transcription, but requires assistance from GTFs for promoter recognition, DNA melting, and TSS selection. A constellation of promoter attributes from DNA sequence to promoter chromatin structure interface with initiation factors to determine initiation output, defined as overall promoter expression level and TSS usage distribution. How these factors control initiation efficiency is not well understood. This dissertation focuses on understanding how DNA sequence, Pol II catalytic activity, and the GTF TFIIH, through its conserved Tfb3 subunit, contribute to TSS selection in

the budding yeast *Saccharomyces cerevisiae*. **Chapter 2** dissects Pol II TSS sequence specificity and how Pol II activity alters it, with **Appendix A** extending investigation to other aspects of promoter architecture. **Chapter 3** examines Tfb3 function in TSS selection and how it genetically interacts with other GTFs.

First, I will review Pol II transcription initiation and the initiation machineries observed in higher eukaryotes and the budding yeast *S. cerevisiae*. Second, I will review the "promoter scanning" model for Pol II initiation in yeast and how this scanning process is controlled by promoter architectural features. Finally, I will review the current understanding of Tfb3/TFIIH contribution to transcription initiation.

1.1 Pol II transcription initiation and TSS selection

1.1.1 Prevalent transcript heterogeneity has significant impacts on gene expression and protein function

Most promoters in eukaryotes specify multiple TSSs, which have been detected by genome-wide approaches (Aanes et al., 2013; Batut et al., 2013; Carninci et al., 2005; Chen et al., 2013; Chia et al., 2021; Consortium et al., 2014; Hoskins et al., 2011; Lu and Lin, 2019; Miura et al., 2006; Nepal et al., 2013; Park et al., 2014; Pelechano et al., 2013; Yamashita et al., 2011; Zhang and Dietrich, 2005; Zheng et al., 2011). For instance, in *S. cerevisiae*, a median of 26 transcript isoforms per coding gene and an average of 7 transcript isoforms per non-coding gene were observed (Pelechano et al., 2013). In mouse, 1.32 5′ TSSs for each 3′ end, with 1.83 3′ ends for each 5′ end, were detected (Carninci et al., 2005). In honeybee, 28–47% of the expressed genes

in the drone head were transcribed from multiple TSSs (Zheng et al., 2011). Together, transcriptional heterogeneity is extensive in eukaryote genomes.

Transcriptional heterogeneity, where transcripts have different lengths and sequences of the 5′ Untranslated Region (UTR), can result in differences in a number of ways. First, transcript isoforms can contain different **post-transcriptional or translational regulatory elements**, such as RNA-binding Protein (RBP) sites and upstream Open Reading Frames (uORF) that can greatly impact post-transcriptional or translational regulation, thereby modulating **protein levels** (Pelechano et al., 2013; Rojas-Duran and Gilbert, 2012). RBP sites have been linked to mRNA expression patterns under different stress conditions (Riordan et al., 2011). uORFs typically serve as inhibitors of translation by restricting ribosome access to the downstream main ORFs (mORFs), though some uORFs can promote mORF translation initiation in response to environmental stresses (Chen et al., 2017; Hinnebusch, 2005; Young and Wek, 2016). Second, alternative TSS usage can drive **protein diversity** via altering **protein function** (Benanti et al., 2009) or **protein localization** (Carlson and Botstein, 1982; Chatton et al., 1988; Natsoulis et al., 1986). One example of altering protein function by alternative TSS usage is the kinesin-associated protein Cik1, which has two developmentally regulated isoforms. During mitosis, Cik1 is expressed as a longer isoform containing an N-terminal domain required for both ubiquitination and nuclear localization and regulates mitotic spindle dynamics. During mating, a shorter Cik1 isoform without the N-terminal domain is expressed and functions in nuclear fusion (Benanti et al., 2009). One example for mRNA composition altering protein localization in yeast is the *SUC2* gene. Its constitutively expressed transcript isoform produces an intracellular invertase, whereas its glucose-regulated transcript contains an additional signal sequence at its 5′ end and therefore produces a secreted form of invertase (Carlson and Botstein, 1982). The mechanism of a single

gene encoding more than one functionally and/or spatially distinct protein isoform via alternative transcript usage can be a very important way for species that do not contain exons and/or rarely use alternative splicing. Third, alternative transcript isoforms can be associated with the **secondary structure** of mRNA, such as hairpin structure within the 5′ UTR (Kozak, 1986) and pseudo-circularized structure of translating mRNA (Christensen et al., 1987), which contribute to translation efficiency. Taken together, the variation in 5′-end of transcripts can lead to different functional consequences in transcriptional and translational regulations.

Because of multiple aspects of functional impacts of alternative TSS usage, TSS usage shifting has been shown to highly relate to gene expression regulation in contexts of cell types (Consortium et al., 2014), developmental processes (Aanes et al., 2013; Batut et al., 2013; Cheng et al., 2018; Chia et al., 2021; Zhang et al., 2017), growth conditions (Lu and Lin, 2019), responses to environmental changes, and cancers (Boyd et al., 2018; Demircioglu et al., 2019; Thorsen et al., 2011). Some examples are the following. In yeast, 45% of core promoters assigned to protein-coding genes were classified as inducible core promoters, defined as their TSS usages were observed to be shifted in response to examined growth conditions (Lu and Lin, 2019). Alternative transcript isoform switching has also been linked to cell-type-specific expression profiles in mice and humans (Consortium et al., 2014), cerebellar development in mice (Zhang et al., 2017), pre- and post-zygotic genome activation in zebrafish (Aanes et al., 2013), and meiotic differentiation in yeast (Cheng et al., 2018). Moreover, TSS usage shifts have also been observed in human tumors and cancers, such as adenoma (Thorsen et al., 2011), pan-cancer (Demircioglu et al., 2019), and inflammatory bowel disease (Boyd et al., 2018). Taken together, shifts in TSS usage are pervasive in various contexts, suggesting the importance of TSS selection in gene expression regulation.

Therefore, the extensive transcript heterogeneity and its considerable impacts on protein production raise an important question: **what mechanism(s) determines why any individual TSS is used?** It should be noted that shifts in TSS usage can be caused by alternative promoter usage and/or alternative TSS usage within one promoter. This dissertation focuses on the latter.

1.1.2 Pol II initiation begins with Pre-initiation Complex (PIC) assembly

Initiation starts with Pre-initiation Complex (PIC) assembly, where Pol II and GTFs are recruited to and assembled at a promoter region, which is a piece of DNA that specifies initiation. PIC formation has long been proposed to be a sequential assembly of factors at a TATA-box (if present) (Buratowski et al., 1989; Fisher et al., 1995; He et al., 2013; Inostroza et al., 1991), though this model has been partially argued against by a recent single-molecule study (Baek et al., 2021) (see later). In the sequential assembly model, TATA-box Binding Protein (TBP), a GTF TFIID subunit, initially recognizes and binds to a TATA-box, followed by the recruitment of TFIIA and TFIIB via direct contact with TBP. Next, Pol II and TFIIF are recruited, followed by recruitment of TFIIE and finally TFIIH. A branched model beyond the simple sequential assembly has been proposed by recent single-molecule microscopy experiments (Baek et al., 2021). Because single-molecule microscopy experiments were performed in the more physiological context of nuclear extract, short-lived intermediates were able to be detected. Interestingly, some Pol II, TFIIF, and TFIIE intermediate complexes have been observed to be pre-assembled at the Upstream Activating Sequence (UAS), independent of the core promoter, and later can be transferred to the core promoter to form a PIC when TBP and TFIIB bind to the core promoter. This observation provided new insight that activators, such as UAS, can facilitate PIC assembly via increasing the local concentration of PIC components during transcription activation and initiation. Then, after PIC

assembly, **how does the PIC determine where and how efficiently transcription initiates, which is known as TSS selection?**

1.1.3 Conserved initiation components support distinct initiation mechanisms in yeast and higher eukaryotes

Although initiation factors and their structures are highly conserved through evolution, TSS selection in higher eukaryotes and yeast are distinct. In higher eukaryotes, transcription initiation for promoters containing TATA-box occurs at a single TSS located 30-31 bp downstream of the TATA-box (Kadonaga, 2012). The precise distance between the TATA-box and TSS suggests that distance is the determinant. In contrast, *S. cerevisiae* initiation for TATA-containing promoters usually occurs at multiple TSSs ranging from 40 to 120 bp downstream of the TATA-box (Qiu et al., 2020; Struhl, 1989; Zhang and Dietrich, 2005). However, yeast PIC assembly occurs at the same location as human relative to the TATA-box (Miller and Hahn, 2006), and promoter DNA melting happens at about 20 bp downstream of the TATA-box, which is comparable to that seen in higher eukaryotes (Buratowski et al., 1991; Giardina et al., 1992; Lis, 1993; Pal et al., 2005). These data suggest that yeast initiation starts 10-90 bp further downstream of PIC assembly position than that in higher eukaryotes. Additionally, the varying TATA-TSS distances suggest distance is not the primary determinant in yeast initiation. Therefore, a "scanning mechanism" for yeast initiation, where a PIC assembles upstream and then scans downstream for TSS selection, was proposed (Kuehner and Brow, 2006; Lis, 1993) and subsequently confirmed as a universal mechanism in the yeast genome (Qiu et al., 2020). This scanning mechanism will be discussed in **Section 1.2**.

The proposed scanning model in yeast immediately raises some interesting questions. First, **how does scanning happen and what factor(s) drive promoter scanning?** TFIIH, or more specifically, Ssl2, the ATP-dependent translocase/helicase of TFIIH, has been proposed to promote DNA melting and drive scanning by translocating DNA into the Pol II active site (Grunberg et al., 2012). TFIIH will be discussed in **Section 1.3**. Second, in addition to TFIIH, **how do other initiation factors contribute to the scanning process?** TFIIE has been suggested to promote scanning processivity by stimulating the ATPase activity of TFIIH (Grunberg et al., 2012; Schilbach et al., 2017) or stabilizing the open complex formation of the PIC (Plaschka et al., 2016). The TFIIH subunit Tfb3 has now been proposed to be the key to couple the promoter scanning ability, driven by TFIIH, to the PIC to support the use of downstream TSSs (unpublished data from my Tfb3 study in **Chapter 3**) (Aibara et al., 2021; Murakami et al., 2015; Schilbach et al., 2021; Schilbach et al., 2017; Yang et al., 2022). Third, **is scanning a conserved mechanism in other eukaryotes**? As mentioned previously, for TATA-containing promoters in higher eukaryotes, TSSs are distance-directed based on the position of the TATA-box. However, only ~8% of human promoters contain a consensus TATA-box (Basehoar et al., 2004). This leads to the provocative question of what mechanism is used by TATA-less promoters to specify TSSs in higher eukaryotes? Most eukaryotic promoters specify multiple TSSs, and most of examined eukaryotes (such as human and zebrafish zygotic promoters) show similar TSS distribution properties (such as TSS spread) as yeast promoters (unpublished data from the Kaplan lab). Our lab has designed experiments that investigate whether perturbation of Pol II activity in *S. pombe* and *Drosophila melanogaster* alters TSS distributions as predicted by the scanning model.

1.2 Yeast Pol II transcription initiation proceeds by promoter scanning and is controlled by promoter architectural features

1.2.1 Promoter scanning model

A wide range of experimental approaches make the budding yeast, *Saccharomyces cerevisiae,* a useful model system for fundamental mechanistic studies. Pol II initiation in yeast proceeds by a proposed promoter scanning mechanism, in which the PIC, comprised of Pol II and GTFs, assembles upstream of the initiation region and then scans downstream to select appropriate TSSs to initiate transcription. Three types of evidence support this scanning model, and genome-wide TSS-seq studies from our lab further suggested it is universal in yeast. First, promoter DNA melting happens at ~20 bp downstream of the TATA-box at *S. cerevisiae GAL1* and *GAL10* promoters (Lis, 1993); however, yeast initiation usually occurs at multiple TSSs, ranging from 40 bp to 120 bp downstream of the TATA-box (Struhl, 1989; Zhang and Dietrich, 2005). These data suggest that initiation in yeast starts 10-90 bp further downstream of the initial DNA melting position and is not distance-directed. Second, polar effects on TSS selection directly supported this unidirectional Pol II scanning model (Kuehner and Brow, 2006). This study showed that Pol II preferred to use the upstream TSS when the promoter contained duplicated TSSs in tandem. Additionally, mutations in upstream TSSs that compromised their usages increased the usage of downstream TSSs, while downstream mutations had no effect on upstream TSSs. Third, Pol II catalytic activity and GTF mutants showed polar effects on TSS distributions at examined promoters and genome-wide by shifting TSS usage upstream or downstream relative WT (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012; Qiu et al., 2020; Zhao et al., 2021), consistent with scanning model predictions. Furthermore, distinct effects on TSS distribution by

different Pol II and GTF mutants further support a "Shooting Gallery" model (see **Section 1.3**). Importantly, recent single molecule studies directly observed scanning (Fazal et al., 2015). Taken together, these observations support that yeast initiation proceeds by a directional promoter scanning mechanism, and this model is a global phenomenon in yeast regardless of promoter class.

Figure 1 shows a simplified view of how the initiation efficiency of any given promoter is predicted to be controlled by architectural features. To understand why any individual TSS is used, I have systematically designed controlled variant libraries dissecting different architectural features to study how TSS usage is affected by or interacts with following promoter architectures: (I) TSS sequence composition to determine Pol II initiation sequence preference, (II) core promoter-TSS distance to assess the role of scanning distance in TSS selection, (III) promoter strength/identity to detect and measure if excess Pol II is present at the downstream edge of initiation windows for different promoters, and (IV) nucleotide composition within the scanning region to determine its effects on TSS usage and expression. Each of these architectural features is discussed in the next section.

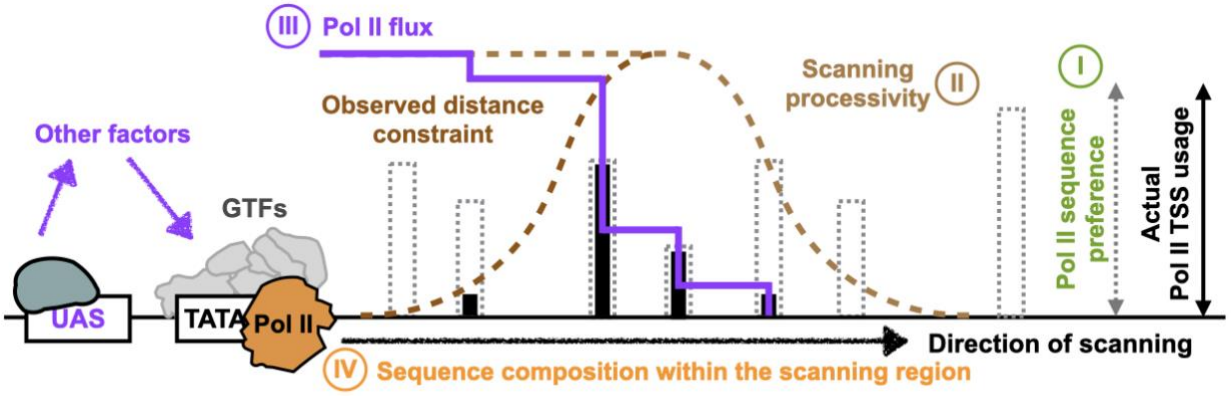


Figure 1 Pol II initiation proceeds by promoter scanning in yeast and is controlled by promoter architectural features

Pol II initiation in yeast proceeds by a promoter scanning mechanism, where the PIC, comprising Pol II and GTFs, assembles upstream of the initiation region and then scans downstream to select appropriate TSSs to initiate transcription.

1.2.2 Contributions of promoter architectural features on TSS selection

1.2.2.1 Pol II TSS sequence specificity

Studies have suggested a $Y_{-1}R_{+1}$ consensus sequence (on the coding strand; Y = pYrimidine at the immediate upstream position of the TSS designated as -1; R = puRine at the TSS position designated as +1) as a near universal transcription initiation motif for RNA polymerase promoters from eukaryotes to phage (Bucher, 1990; Carninci et al., 2006; Chen et al., 2013; Corden et al., 1980; Cortes et al., 2013; Furter-Graves and Hall, 1990; Hahn S, 1985; Hashimoto et al., 2004; Healy et al., 1987; Kim et al., 2012; Kuehner and Brow, 2006; Lu and Lin, 2019; Malabat et al., 2015; McNeil and Smith, 1985; Policastro et al., 2020; Qiu et al., 2020; Smale and Baltimore, 1989; Suzuki et al., 2001; Vvedenskaya et al., 2015b; Zhang and Dietrich, 2005; Zheng et al., 2011). This $Y_{-1}R_{+1}$ sequence preference is likely due to the stacking of an incoming purine NTP and a purine base at position -1 on the template strand (meaning a pyrimidine at position -1 on the transcribed strand) (Basu et al., 2014; Gleghorn et al., 2011). Furthermore, genome-wide TSS data revealed a strong A (Adenine) preference at position -8 relative to the TSS in *S. cerevisiae* (referred to as A_{-8} or -8A throughout) (Lu and Lin, 2019; Malabat et al., 2015; Policastro et al., 2020; Qiu et al., 2020; Zhang and Dietrich, 2005). In addition, promoters with high expression and focused initiation at mostly a single TSS tend to have additional A enrichment at positions -7 to -3 (Lu and Lin, 2019; Lubliner et al., 2013; Maicas and Friesen, 1990; Qiu et al., 2020; Zhang and Dietrich, 2005). All these observations indicate that the TSS sequence and its sequence context are primary

determinants for the initiation output; however, we lack a detailed understanding of **what sequences make efficient TSSs**.

Existing promoters in the genome have limited context and are regulated by diverse architectural features, such as chromatin context and core promoter-TSS distance. These confounding factors, together with biased base distribution shaped during evolution, obscure our ability to understand TSS sequence contribution to initiation by analyzing genomic data. For example, genome-wide TSS mapping data from our lab demonstrated that Pol II activity mutants conferred changes to TSS selection, including apparently altered TSS motif usage (**Figure 2A**) and shifted TSS distribution (**Figure 2B**) (Qiu et al., 2020). More catalytically active Pol II mutants (termed GOF for "Gain-of-Function"), which shifted TSSs upstream, showed overall decreased selectivity for -8A. Conversely, Pol II reduced activity mutants (termed LOF for "Loss-of-Function"), shifting TSSs downstream, showed apparent increased preference for -8A. This raises an intriguing question: **what is the relationship between Pol II catalytic activity and TSS motif usage?** It is unclear if the observed TSS shifts are a consequence of altered TSS motif preference or vice versa. One simple model is that Pol II mutants altered catalytic efficiency for all sequence motifs, and the apparent altered sequence specificity resulted from the following two causes (Model 1 in **Figure 2C**). First, given the observations that $A_{-8}Y_{-1}R_{+1}$ motifs are mainly used, and TSS-TSS distances are enriched in 8 nt in yeast (Pelechano et al., 2013), and with an ideal assumption of evenly distributed bases in the genome, 50% A_{-8} would have a Y_{-9}, which means A_{-8} could also function as an A_{+1} for an upstream TSS (**Figure 2D**). This further predicts that any increase in overall efficiency would increase the usage of the upstream TSS, which would only be an $A_{-8}Y_{-1}A_{+1}$ motif, resulting in a motif ~25% of the time, if base distributions were even among bases. This would result in a usage shift from $A_{-8}Y_{-1}R_{+1}$ to $B_{-8}Y_{-1}R_{+1}$ (B = not A). Second, unevenly

distributed TSS motifs are a confounding variable for apparent alterations to sequence specificity. This is illustrated by the asymmetric localization of $A/B_{-8}Y_{-1}R_{+1}$ motifs in yeast promoters around TSS regions (Qiu et al., 2020), where preferred and less preferred TSS motifs showed enrichment downstream or upstream of median TSSs, respectively. An alternative model is that Pol II mutants changed the specificity for TSS motifs, or even more specifically, for the -8A, resulting in shifts of TSS utilization (Model 2 in **Figure 2C**). Since promoters are evolutionary products, it is difficult to distinguish these two models by studying existing gene promoters *in vivo*, because yeast promoters have context limitations and are affected by regulatory elements other than TSS sequence. To sidestep potential confounding variables and specifically dissect Pol II TSS sequence specificity and how Pol II activity alters this specificity, I have developed a massively parallel report assay "Pol II MASTER" (Pol II MAssively Systematic Transcript End Readout) to investigate the initiation efficiency of ~1 million unique TSS sequences in WT and Pol II catalytic mutants (see **Chapter 2**).

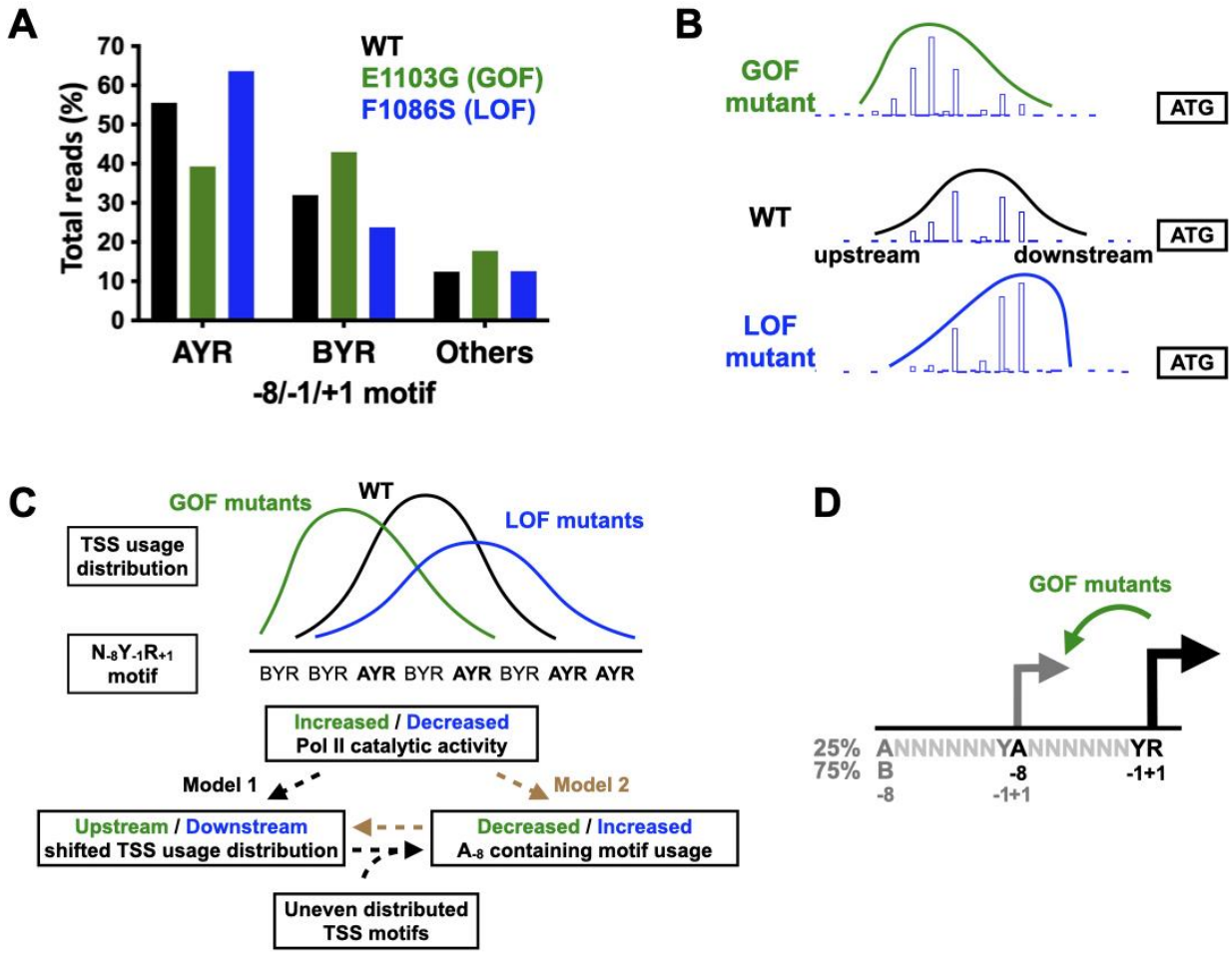


Figure 2 Pol II activity mutants effects on TSS selection observed in genomic studies

(A) Pol II activity mutants showed apparent altered TSS usage levels for $A_{-8}Y_{-1}R_{+1}$ and $B_{-8}Y_{-1}R_{+1}$ motifs. Y = pYrimidine (C or T); R = puRine (A or G). **(B)** Pol II activity mutants showed polar effects on TSS distribution. Pol II GOF mutants shifted TSS usage distribution upstream relative to WT distribution, whereas LOF mutants shifted TSS usage distribution downstream. **(C)** Two potential models of the causality between TSS distribution shifts and TSS motif preference changes in Pol II activity mutants. **(D)** Upstream shifting of TSS usage distribution would cause decreased TSS usage of $A_{-8}Y_{-1}R_{+1}$ motif.

1.2.2.2 Core promoter-TSS distance constraints

Genome-wide mapping data from our lab and other groups suggest that the distance between the core promoter for genes with TATA-box and TSSs in yeast is 40-120 bp (Qiu et al., 2020; Zhang and Dietrich, 2005). Though there is no fixed distance relationship between TATA-box and TSS like that found in higher eukaryotes, there are limitations for distance in which TSSs can be used. When TSSs are located too close or too far away from the TATA-box or the assembly position, they are not able to be efficiently used or may even not be recognized. Deletion and insertion studies have learned limitations of a "scanning window". First, the lower limit of the scanning window was determined as about 50-60 bp from the PIC assembly position (i.e., TATA-box or other) (Faitar et al., 2001; Hahn S, 1985; Nagawa and Fink, 1985), but the mechanism for this is not known. Deletion experiments for yeast promoters observed that shortened TATA-TSS distance could cause a significant reduction in TSS efficiency (Faitar et al., 2001; Hahn S, 1985), and a 44 bp TATA-TSS distance caused a TSS not to be recognized (Nagawa and Fink, 1985). This minimal distance requirement might be required for open complex formation, suggested by an open complex structure model with 34 nt DNA connecting the upstream TATA-box and the TSS (Kostrewa et al., 2009). Second, an upper limit of the scanning window has been suggested as about 110-120 bp from the PIC assembly position (Hahn S, 1985; Nagawa and Fink, 1985), although PICs may reach further downstream TSSs at some promoters, such as *IMD2* (Jenks et al., 2008). Detectable promoter melting was found to extend from 20 bp to 120 bp downstream of the TATA-box at *GAL1* and *GAL10*, suggesting a limit to the scanning process (Lis, 1993). The processivity of TFIIH, the proposed scanning energy provider and engine for DNA translation, has been predicted to biochemically limit the scanning distance. Importantly, the processivity of TFIIH within the PIC has been measured to be 94 ± 36 bp (mean ± s.d.) by real-time single molecule

studies of downstream DNA compaction (downstream DNA moving closer to upstream DNA) (Fazal et al., 2015). Based on these observations, we speculate that promoters in yeast may have two kinds of distance windows: a "scanning window" in which TSSs are reachable for Pol II and an "efficiency window" in which TSSs can be efficiently used (**Figure 3**).

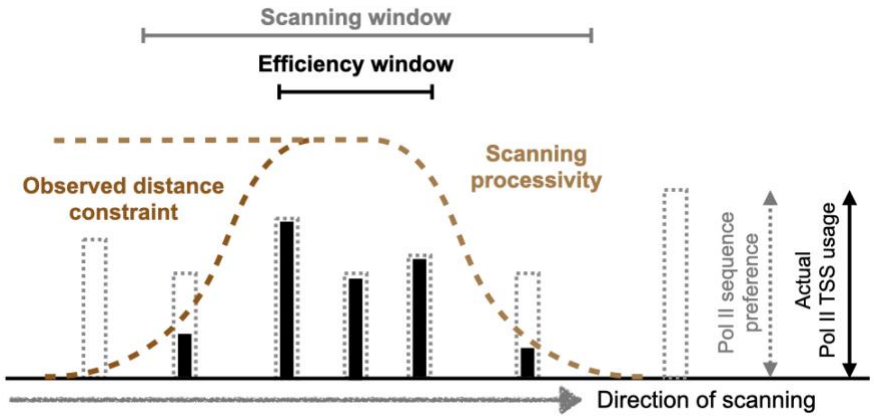


Figure 3 Definition of scanning window and efficiency window

Two kinds of distance windows contribute to TSS selection. "Scanning window" represents the region in which TSSs could be used by Pol II. "Efficiency window" represents the region in which TSSs could be efficiently used.

1.2.2.3 Promoter strength

Upstream Activation Sequences (UASs), as cis-acting regulatory elements, are usually positioned 250-400 bp upstream of TSSs in yeast and direct transcription initiation events. Two interesting questions about promoter identity are (1) **how efficiently do native promoters convert recruited Pol II into initiation** and (2) **do UASs also regulate TSS distribution in addition to overall expression level?**

A study from the Segal lab has suggested that changes to core promoter sequence can increase expression level (Lubliner et al., 2015), indicating that native promoters might waste some Pol II flux (the amount of Pol II recruited to the promoter) if the processivity limit is reached before

encountering a strong TSS (**Figure 4A**). This hypothesis could be tested by inserting a strong TSS downstream to investigate whether it would be used (**Figure 4B**) (see **Appendix A**).

If initiation is entirely efficient, or for those promoters that efficiently initiate all recruited Pol II, are the transcriptional capacity and TSS distribution of a particular core promoter limited by its coupling UAS (**Figure 4C**)? Synthetic hybrid and artificial promoter studies have shown that different pairs of UASs and core promoters can positively and negatively affect expression levels (Blazeck et al., 2012; Dhillon et al., 2020; Liu et al., 2020), suggesting that there are some possible functional interactions between UASs and core promoters. Moreover, when a UAS was placed far away from a core promoter, it was shown to be able to initiate transcription proximally in the absence of nearby core promoter elements (Dobi and Winston, 2007). This suggests that directing transcription initiation at proximal sites may be an inherent property of UASs. These observations raise the possibility that promoter identity may contribute to TSS distribution in addition to overall expression level. Specifically, Pol II flux determined by promoter identity might limit the downstream edge of the scanning window, which contains TSSs that have the chance to be reached (**Figure 4C)**. If this is the case, where Pol II flux is a determinant for TSS distribution, then coupling a particular core promoter with different UASs would alter TSS distribution shapes instead of simply changing overall usage (**Figure 4D)** (see **Appendix A**).

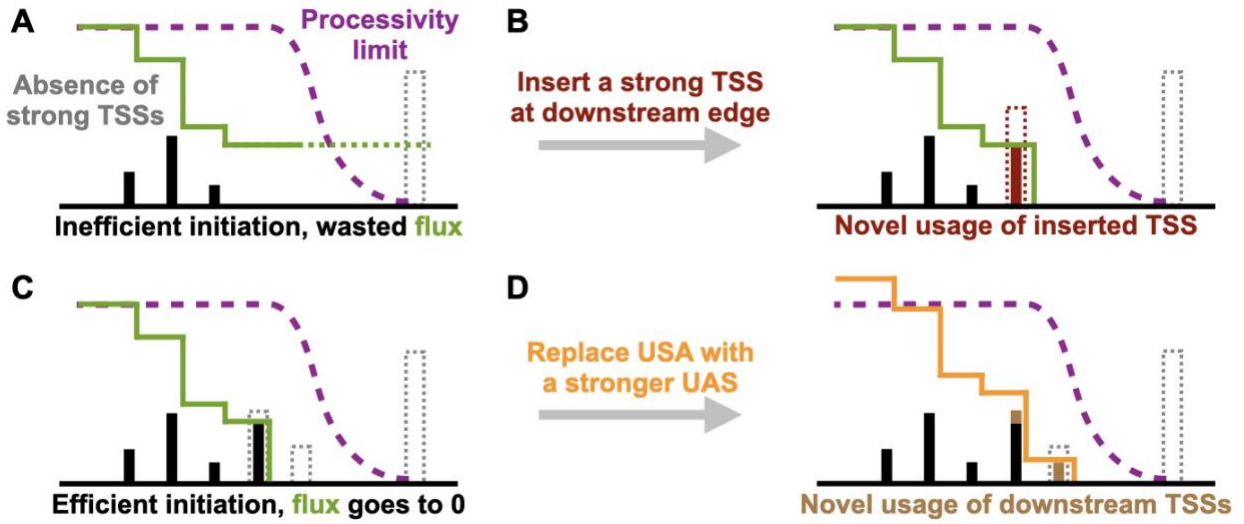


Figure 4 Proposed Pol II flux effects on TSS selection

(A) Native promoters might waste some Pol II flux if the processivity limit is reached before encountering a strong TSS. **(B)** Wasted Pol II flux may be detected by insertion of a strong TSS at the downstream edge of native TSS usage distribution. If novel usage of the inserted TSS can be detected, it suggests the examined promoter originally wasted some Pol II flux. **(C)** UAS might determine the downstream edge of the scanning window by determining Pol II flux. **(D)** UAS contribution on TSS selection may be detected by replacing native UAS with a stronger UAS. If increased usage of the TSS at the downstream edge or novel usage of further downstream TSS(s) can be detected, it suggests UASs regulate TSS distribution in addition to overall expression level.

1.2.2.4 Scanning region sequence composition

The scanning region is defined as the region between the PIC assembly position (TATA-box, if present) and TSSs. Yeast genomic and mutational studies have shown that promoters with higher activity tend to have a pyrimidine-rich scanning region on the coding strand, especially T-richness (Lubliner et al., 2013; Lubliner et al., 2015; Maicas and Friesen, 1990; Wu and Li, 2010). This enrichment might be for many potential reasons. First, T-richness might provide lower +1 nucleosome occupancy, therefore supporting higher promoter activity, because G/C content is suggested to be highly related with intrinsic nucleosome occupancy (Lee et al., 2007; Peckham et

al., 2007; Segal et al., 2006; Tillo and Hughes, 2009). Yet, it should be noted that technical artifacts and MNase sequence selectivity might also contribute to distinctions between A-T rich DNA and G-C rich DNA. Second, T:A pair base enrichment, namely T-richness on the coding strand and A-richness on the template strand, may be helpful for double-stranded DNA melting and therefore promote transcription bubble formation (Bansal et al., 2014). Though either T:A or A:T pair base could provide an easily meltable region, the strand bias (T on the coding strand and A on the template strand) might be about controlling initiation, meaning transcription can initiate at A but not T generally. In line with this, the bias for another pyrimidine, C, on the coding strand might also minimize initiation upstream. Third, pyrimidine-richness, especially T-richness, can help to accumulate functional mRNAs by inhibiting upstream Cryptic Unstable Transcripts (CUTs). Some TATA-box proximal T-rich regulatory sequences have been shown to promote transcription termination/degradation of the upstream CUTs for some yeast genes, such as *CYC1* (McNeil, 1988) and *URA2* (Thiebaut et al., 2008). The degradation of upstream CUTs helps ensure only functional transcripts initiating from downstream appropriate TSSs would be accumulated. Altogether, nucleotide composition within the scanning region may be expected to contribute to TSS distribution in different directions.

Sequence composition within the promoter region may have complex relationships to expression. For example, high T% or T-rich elements might be necessary but not sufficient for high promoter initiation output. T content or motifs might also compete with other base features, such as G/C percentage. Artificial promoter studies have shown that a promoter with higher T% and more T-rich elements showed lower expression level than another promoter containing more G and poly G (Liu et al., 2020), suggesting sequences within the scanning region have higher order influences on initiation. In order to test this, a comprehensive study investigating how T content,

sequence order, and potential sequence motifs within the scanning region contribute to initiation output, including both TSS selection and overall expression level, is needed (see **Appendix A**).

1.2.3 "Shooting Gallery" model for initiation by promoter scanning and contributions of transcription factors to TSS selection

A number of mutations conferring polar effects on TSS distribution at tested promoters or genome-wide have been identified in Pol II (Rpb1, Rpb2, Rpb7, Rpb9), GTFs TFIIB, TFIIH (Ssl2, Tfb3), TFIIE (Tfa1), TFIIF (Tfg1, Tfg2), and the transcriptional coactivator Sub1 (Berroteran et al., 1994; Braberg et al., 2013; Chen and Hampsey, 2004; Faitar et al., 2001; Freire-Picos et al., 2005; Ghazy et al., 2004; Goel et al., 2012; Jin and Kaplan, 2014; Kaplan et al., 2012; Khaperskyy et al., 2008; Kuehner and Brow, 2006; Majovski et al., 2005; Pardee et al., 1998; Pinto et al., 1992; Pinto, 1994; Qiu et al., 2020; Sun et al., 1996; Wu et al., 1999; Zhao et al., 2021) (unpublished Tfb3 data in **Chapter 3**). As described earlier, the polar effects on TSS distribution support the promoter scanning model for yeast initiation. Furthermore, distinct patterns of polar shifts in TSS distribution by different Pol II and GTF mutants and genetic interactions between factors (see later) further support a "Shooting Gallery" model (**Figure 5**). This model imagines the Pol II active site as a fixed firing position while considering potential TSS(s) on the DNA template being scanned as target(s) moving relative to the Pol II active site. Therefore, the Shooting Gallery model predicts that the rate of firing and the rate of a target passing together determine the probability a passing target is hit. Additionally, the processivity of targets passing (DNA scanning) determines how many targets (how many DNA bases) have the potential to pass the fixed firing position (Pol II active site). Here, Pol II catalytic activity determines how fast the first phosphodiester bond is catalyzed (i.e., the rate of firing), and TFIIH determines both how fast (i.e., the rate at which a

target passes) and how far (i.e., the potential of targets passing the firing position) the DNA template is inserted into the Pol II active site. Therefore, the critical prediction of this Shooting Gallery model is that TSS distribution is controlled by both Pol II catalytic activity, controlling RNA synthesis, and TFIIH (or more specifically, the ATP-dependent translocase/helicase subunit of TFIIH, Ssl2) enzymatic activity, controlling DNA scanning. Further, initiation factors that are in coordination with these two determinants, Pol II and TFIIH, are assigned to two major functional networks designated as "efficiency" and "processivity" networks, respectively. Based on studies from our lab and others, Pol II, TFIIB, and TFIIF likely function within the efficiency network, and TFIIH (Ssl2) and Sub1 function within the processivity network. Each of the initiation factors involved in TSS selection is discussed below.

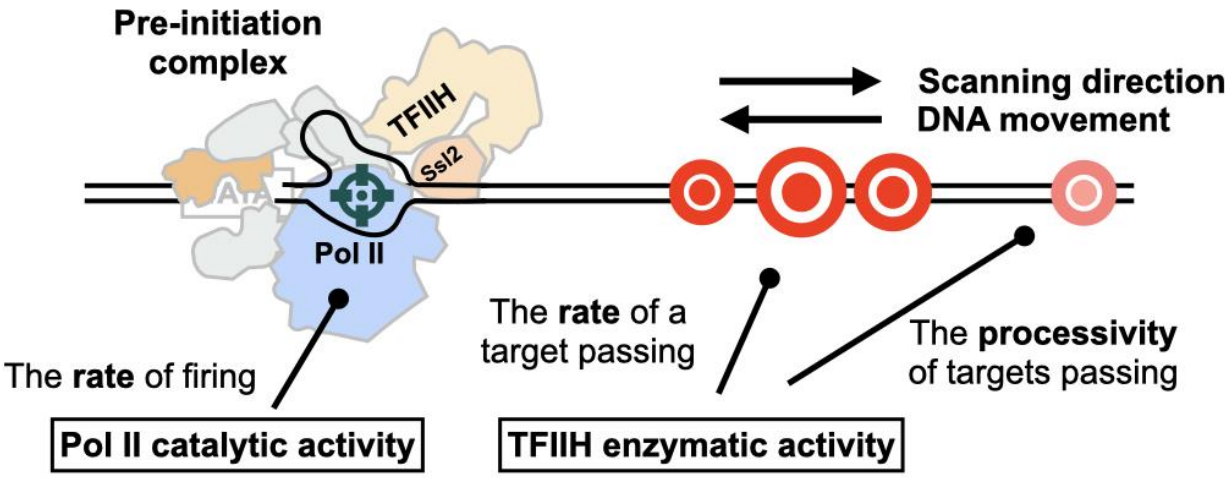


Figure 5 "Shooting Gallery" model for Pol II initiation by promoter scanning in yeast

Pol II initiation in yeast proceeds by promoter scanning, in which PIC assembles upstream of the initiation region and then scans downstream to select appropriate TSSs to initiate transcription. The Shooting Gallery model predicts that TSS distribution is controlled by both Pol II catalytic activity, determining how fast the first phosphodiester bond is catalyzed (i.e., the rate of firing), and TFIIH/Ssl2 enzymatic activity, controlling both how fast (i.e., the rate of a target passing) and how far (i.e., the processivity of targets passing) the DNA template is inserted into the Pol II active site.

1.2.3.1 Pol II

Yeast Pol II comprises 12 highly conserved subunits, Rpb1-Rpb12. A number of mutants that alter TSS distribution have been identified in Pol II subunits (Rpb1, Rpb2, Rpb7, Rpb9). Although the mechanism of the alternation is not fully understood, extensive studies have suggested that mutations mainly either affect Pol II catalytic activity that is coordinated by the Rpb1 active site "Trigger Loop" (TL, Rpb1 residues 1076-1106) or affect GTF functions through altering corresponding interfaces between Pol II and GTF(s).

Mutations in the Rpb1 could alter TSS selection in both directions, namely shifting TSS distribution upstream or downstream (Berroteran et al., 1994; Braberg et al., 2013; Kaplan et al., 2012; Majovski et al., 2005; Thiebaut et al., 2008). Alleles with mutations in the TL have been further shown to alter *in vitro* elongation rates. One class of mutants shifting TSS usage upstream show increased catalytic rate *in vitro* and therefore are termed "Gain-of-Function" (GOF). Another class of mutants shifting TSS usage downstream show decreased catalytic rate and are termed "Loss-of-Function" (LOF). In addition, Pol II and TFIIB or TFIIF alleles showed broad additive or suppressive interactions on TSS alteration, suggesting they are in the same pathway (the efficiency network).

All identified alleles of *rpb2, rpb7,* and *rpb9* generate upstream shifts of TSS usage but likely via different mechanisms (Braberg et al., 2013; Chen and Hampsey, 2004; Chen et al., 2007; Furter-Graves et al., 1994; Hull et al., 1995; Khaperskyy et al., 2008; Sun et al., 1996). Structural and genetic studies have revealed that Rpb2 and Rpb9 function in TSS selection by affecting the conformation and/or activity of TFIIF (Chen et al., 2007; Khaperskyy et al., 2008), mutations in which also confer TSS upstream shifting (Freire-Picos et al., 2005; Ghazy et al., 2004; Jin and Kaplan, 2014; Qiu et al., 2020). In addition, Rpb9 has been suggested to indirectly affect the

mobility of Pol II TL (Kaster et al., 2016; Walmacq et al., 2009). The Pol II stalk subunit Rpb7 has recently been proposed to function in TSS selection by modulating the association between Pol II and TFIIH via the Pol II stalk-TFIIH-TFIIE interface (see **Chapter 3**) (Schilbach et al., 2021; Schilbach et al., 2017).

1.2.3.2 TFIIB

Yeast TFIIB is encoded by the *SUA7* gene, mutations in which were first identified as suppressors of an aberrant ATG translation start codon in the *cyc1-5000* mutant (Pinto et al., 1992). As a GTF, TFIIB has been demonstrated to have pleiotropic roles during transcription initiation, including PIC assembly, promoter opening, TSS selection, initial RNA synthesis, and initiation-elongation transition, through its different domains ("B-ribbon", "B-reader", "B-linker", "B-core", and C-terminal tail). First, during PIC assembly, TFIIB recruits Pol II to the promoter via interactions of its B-ribbon domain with Pol II and of its B-core domain with DNA and TBP (Bushnell DA, 2005; Chen and Hahn, 2003; Kostrewa et al., 2009; Nikolov et al., 1995). Second, during promoter opening, the B-linker domain of TFIIB binds to the Pol II rudder and clamp coiled-coil and helps to position DNA, therefore assisting promoter opening (Kostrewa et al., 2009; Sainsbury et al., 2013). Third, and importantly, the TFIIB B-reader domain binds to the DNA template strand upstream to assist TSS selection during promoter scanning. Specifically, the TFIIB B-reader domain has been proposed to function as an anchor point to pause the scanning process and therefore promote Pol II initiation. This proposed role of TFIIB has been supported by multiple observations from my promoter libraries (**Chapter 2**) and other groups (Bangur et al., 1997; Chen and Hampsey, 2004; Faitar et al., 2001; Jin and Kaplan, 2014; Kuehner and Brow, 2006; Pardee et al., 1998; Pinto, 1994; Sainsbury et al., 2013; Yang and Ponticelli, 2012). Structural studies have observed the direct contact of TFIIB B-reader residues R64 and D69 and -

7T and -8T on the template strand in a Pol II-TFIIB complex structure (Sainsbury et al., 2013). My massive promoter libraries detected sequence interaction between positions -8 and -7 (see **Chapter 2**). Mutagenesis work has shown that mutations or alanine insertions in the B-reader domain (D58, E62, W63, R64, F66, N68, D70, P76, R78, and V79) caused downstream TSS distribution shifts (Bangur et al., 1997; Chen and Hampsey, 2004; Faitar et al., 2001; Jin and Kaplan, 2014; Kuehner and Brow, 2006; Pardee et al., 1998; Pinto, 1994; Sainsbury et al., 2013; Wu et al., 1999; Yang and Ponticelli, 2012). Some of these mutations additionally showed different sensitivities to different TSS sequences and sequence contexts (Faitar et al., 2001). In contrast to mutations in the B-reader domain, all examined mutations in other domains, with one exception of C149Y, showed no effect on TSS selection (Bangur et al., 1997; Wu et al., 1999), suggesting that TFIIB may regulate TSS selection specifically via its B-reader domain. Fourth, after initiation, TFIIB stimulates initial RNA synthesis by a proposed allosteric mechanism of rearranging active site residues in cooperation with TFIIF, stabilizing a closed polymerase clamp, preventing tilting of short DNA-RNA hybrids, and subsequently separating growing RNA (>6 nt) from DNA template and directing RNA to Pol II RNA exit tunnel (Cabart et al., 2014; Sainsbury et al., 2013). Fifth, TFIIB may contribute to promoter escape by being released from the initially transcribing complex and therefore enabling elongation complex formation, when the RNA length reaches 12-13 nt (Bushnell DA, 2005; Sainsbury et al., 2013). Nevertheless, the TFIIB release may be not sufficient for promoter escape, because TFIIB can be associated with complexes when the nascent RNA reaches even 49 nt, where promoter escape is often completed (Fujiwara et al., 2019). In addition, TFIIB, together with Pol II stalk subcomplex Rpb4-Rpb7, has also been suggested to function in the recruitment of transcriptional co-activator Sub1 to Pol II (Garavis et al., 2017).

1.2.3.3 TFIIE

Yeast TFIIE is a heterodimer, composed of Tfa1 (TFIIEα in humans) and Tfa2 (TFIIEβ in humans) (Feaver et al., 1994), both of which are essential for viability. TFIIE has been suggested to contribute to initiation in multiple ways via contacts with TFIIH, TFIIF, Pol II, and DNA. First, TFIIE is required for TFIIH recruitment to the PIC via interactions between TFIIE subunit Tfa1/TFIIHα (in yeast/humans) and TFIIH subunits Ssl2/XPB and Tfb1/p62 (Maxon et al., 1994; Ohkuma et al., 1995; Okuda et al., 2004; Okuda et al., 2008; Schilbach et al., 2017). Second, TFIIE contributes to promoter DNA opening by positioning Ssl2 downstream of the PIC, capturing non-template strand promoter DNA generated by Ssl2 via its tandem WH domains (Grunberg et al., 2012), and likely affecting the conformational ratcheting of Ssl2 ATPase during DNA translocation via the contact between the Tfa1 E-bridge and Ssl2 lobe (Schilbach et al., 2021; Schilbach et al., 2017). Third, TFIIE and TFIIF together stabilize the open complex of the PIC by binding open promoter DNA from opposite sides of the Pol II cleft (Plaschka et al., 2016). Fourth, TFIIE helps to insert the template strand into the Pol II cleft. Fifth, TFIIE might promote scanning processivity via its interface with TFIIH subunit Tfb3 (unpublished *TFA1* and *TFB3* screen results from the Kaplan lab) (Schilbach et al., 2021). Importantly, *TFA1* screen from our lab (unpublished) has identified mutations that can confer either upstream or downstream shifting of TSS distribution, supporting the role of TFIIE in TSS selection though the mechanism is not fully understood.

1.2.3.4 TFIIF

Yeast TFIIF complex is comprised of three subunits, including two essential mammalian homologs Tfg1 and Tfg2 and one yeast-only subunit Tfg3 (Henry et al., 1994). Although all three

TFIIF subunits have been shown to participate in transcription (Henry et al., 1994), Tfg3, also known as Taf30, is also a component of TFIID (GTF), Mediator (transcriptional coactivator), and SWI/SNF (chromatin remodeling) complexes (Cairns et al., 1996; Henry et al., 1994; Kim et al., 1994), suggesting it likely functions in transcription in ways distinct from Tfg1 and Tfg2. TFIIF has been demonstrated to contribute to TSS selection and early elongation, via direct interaction with Pol II and functional interactions with TFIIB (Chen et al., 2007; Ghazy et al., 2004; Jin and Kaplan, 2014; Khaperskyy et al., 2008). Mutations in Tfg1 and Tfg2 (*tfg1* E346A, *tfg1* W350A, *tfg1* G363D, *tfg1* G363E, *tfg2* L59K, *tfg2*Δ*146-180*, *tfg2*Δ*261-273*) have been shown to confer upstream shifting of TSS distribution, establishing a role for TFIIF in TSS selection (Freire-Picos et al., 2005; Ghazy et al., 2004; Jin and Kaplan, 2014; Qiu et al., 2020). Further, analyses of double mutants demonstrated a wide range of genetic interactions between TFIIF (Tfg1, Tfg2), Pol II (Rpb1, Rpb9), and TFIIB (Freire-Picos et al., 2005; Ghazy et al., 2004; Jin and Kaplan, 2014), suggesting these factors appear to function in TSS selection through the efficiency network. Moreover, TFIIF has been proposed to stimulate early phosphodiester bond formation and stabilize a short RNA-DNA hybrid in the Pol II active center via direct interaction with Pol II Rpb2 lobe and Rpb9 (Cabart et al., 2014; Chen et al., 2007; Khaperskyy et al., 2008).

1.2.3.5 TFIIH

Yeast TFIIH is a multiprotein complex of 11 subunits (Rad3, Tfb1, Tfb2, Tfb3, Tfb4, Tfb5, Ssl1, Ssl2, Kin28, Ccl1, and Tfb6) (Grunberg and Hahn, 2013; Murakami et al., 2012; Ranish et al., 2004; Svejstrup et al., 1994). As described in the Shooting Gallery model, TFIIH, or more specifically, Ssl2, is predicted to determine both how fast and how far the DNA template is inserted into the Pol II active site. Furthermore, this model hypothesizes that two distinct classes of *ssl2* mutants may exist. First, *ssl2* "rate" mutants altering the rate of scanning are predicted to show

similar phenotypes on TSS distributions as Pol II efficiency mutants. Second, *ssl2* "processivity" mutants altering scanning window/distance are predicted to show distinct phenotypes from Pol II mutants. These predictions are supported by different patterns of TSS usage at the *ADH1* promoter for specific classes of Pol II and *ssl2* mutants (Kaplan et al., 2012; Zhao et al., 2021). Unlike Pol II upstream shifting mutants where upstream TSSs that were never or rarely used in WT were used, examined *ssl2* mutants shifted TSSs upstream by increasing the usage of existing upstream TSSs. This observation was consistent with the Shooting Gallery model in which the Pol II catalytic activity and TFIIH processivity contribute to TSS distribution in parallel. Further, a recent study from our lab investigating genetic interactions between Ssl2 and Pol II or GTF alleles has revealed that Sub1, but not TFIIB or TFIIF, functions in TSS selection within the processivity network (Zhao et al., 2021). In addition, results of my *TFB3* screen (**Chapter 3**) suggest that Tfb3, together with the Pol II stalk subunit Rpb7, might also function through the processivity network.

1.2.3.6 Sub1, a transcriptional co-activator

Yeast Sub1, homolog of human coactivator PC4, was originally isolated as a suppressor of two TFIIB mutants (E62G and R78H) (Knaus et al., 1996). Sub1 has been implicated in different stages of transcription, including TSS selection (Braberg et al., 2013; Koyama et al., 2008), initiation-elongation transition by modulating Pol II CTD phosphorylation (Garavis et al., 2017), transcription elongation rate via association with Spt5 (Garcia et al., 2012), mRNA splicing (Braberg et al., 2013), and mRNA 3′-end processing (He et al., 2003). Sub1 has been identified as a PIC component (Sikorski et al., 2011) and was proposed to physically interact with the junction between single- and double-stranded DNA (Sikorski et al., 2011), Pol II stalk subcomplex Rpb4-Rpb7 (Garavis et al., 2017), TFIIB (Knaus et al., 1996), and TFIID subunit TBP (Knaus et al., 1996). Importantly, deletion of *SUB1* (*sub1*Δ) caused a significant downstream shifting of TSS

distribution (Braberg et al., 2013; Koyama et al., 2008), suggesting an important role of Sub1 during transcription initiation. Additionally, genetic interactions of Sub1 with other initiation factors (Pol II, Ssl2) further suggested that Sub1 might be a negative factor of TFIIH/Ssl2 processivity. First, *sub1*Δ showed an epistatic relationship with Pol II GOF mutants (L1101S and E1103G) and synthetic sickness with Pol II LOF mutants (N1082S, H1085Q, and F1086S) (Braberg et al., 2013). This genetic interaction pattern is distinct from how Pol II alleles interact with TFIIB or TFIIF alleles (Jin and Kaplan, 2014), suggesting Sub1 might be involved in an alternative pathway from the Pol II-TFIIB-TFIIF network. Second, and importantly, a recent Ssl2 study from our lab has shown that *sub1*Δ effects on TSS distribution require WT Ssl2 function (Zhao et al., 2021), suggesting Sub1 contributes to TSS selection through regulating TFIIH Ssl2 processivity.

1.3 Tfb3, a TFIIH subunit, bridges between Pol II, TFIIE, and other TFIIH subunits

1.3.1 Overview of TFIIH and its subunits

Yeast TFIIH is a multiprotein complex of 11 subunits, including ten human homologs (Rad3, Tfb1, Tfb2, Tfb3, Tfb4, Tfb5, Ssl1, Ssl2, Kin28, and Ccl1) (Grunberg and Hahn, 2013; Ranish et al., 2004; Svejstrup et al., 1994) and one yeast specific subunit Tfb6 (Murakami et al., 2012). TFIIH is involved in multiple fundamental processes including RNA transcription, Nucleotide Excision Repair (NER) of DNA damage, and cell cycle control, via the complete TFIIH protein complex (termed "holoTFIIH") or various functional subcomplexes ("core" TFIIH and "TFIIK" module). First, the holoTFIIH complex is required for RNA Pol II transcription and is

the only GTF possessing catalytic activity (the kinase Kin28, the ATPase and ATP-dependent translocase Ssl2, and the 5′ to 3′ DNA helicase Rad3). During transcription, TFIIH promotes DNA unwinding, drives promoter scanning, and phosphorylates Pol II C-Terminal Domain (CTD) to stimulate the transition between initiation and elongation. Second, the seven-subunit subcomplex core TFIIH, including Rad3 (XPD in humans), Tfb1 (p62), Tfb2 (p52), Tfb4 (p34), Tfb5 (p8), Ssl1 (p44), and Ssl2 (XPB), is required for NER. During NER, the core TFIIH opens damaged DNA via Ssl2/XPB ATPase activity, allowing Rad3/XPD to track on strand via its helicase activity to verify the damage for nucleases to excise (Compe and Egly, 2012). Third, the three-subunit kinase module TFIIK (counterpart of human CDK), containing kinase Kin28 (CDK7 in humans), cyclin Ccl1 (Cyclin H), and Tfb3 (MAT1), is required for phosphorylation of the CTD of Pol II Rpb1 during transcription but is dissociated from the core TFIIH during NER.

1.3.2 Tfb3 domains and interactions with other initiation factors

Tfb3 (MAT1 in humans) is a 38 kDa molecular mass protein and is essential for cell viability (Faye et al., 1997; Feaver et al., 1997). Structures of Tfb3 and its homologs in other species have been reported at different resolutions (representatives are summarized in **Table 1**). Tfb3 comprises an N-terminal RING finger domain (residues 1-69), an ARCH anchor domain (residues 70-145), an α-helical domain (residues 146-236), and a C-terminal hydrophobic region (residues 237-321) (**Figure 6**).

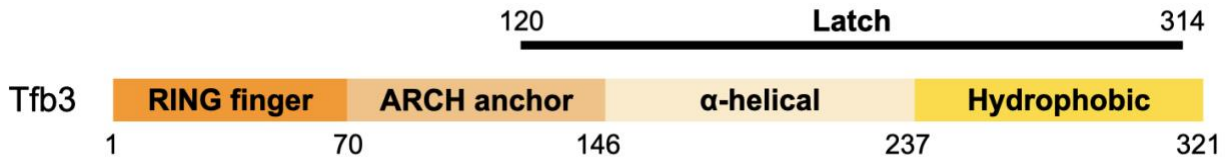


Figure 6 Schematic diagram of domains and structural motifs of Tfb3

Residue numbers are given for domain or motif borders.

Table 1 Representitive structure studies involving Tfb3 and its homologs

Ref.	Title	Description	Species	Key PDB IDs	Year	Tfb3/MAT1 region contained in structure
(Schilbach et al., 2017)	Structures of transcription pre-initiation complex with TFIIH and Mediator	Structures show Tfb3 RING-Pol II stalk-TFIIE Tfa1 interface and TFIIH-Pol II-Mediator interface.	*S. cerevisiae*	5OQJ (PIC), 5OQM (PIC–cMed complex)	2017	8-145 (full length=321)
(Greber et al., 2017)	The cryo-electron microscopy structure of human transcription factor IIH	Structure of human TFIIH in free form	*H. sapiens*	5OF4	2017	45-168 (full 309)
(Greber et al., 2019)	The complete structure of the human TFIIH core complex	Structure of human TFIIH in free form	*H. sapiens*	6NMI	2019	1-210 (full 309)
(van Eeuwen et al., 2021)	Structure of TFIIK for phosphorylation of CTD of RNA polymerase II	Structure of yeast TFIIK (Kin28/Ccl1/Tfb3) complex. The C-terminal hydrophobic of Tfb3 is resolved.	*S. cerevisiae*	7KUE	2021	259-320 (full 321)
(Schilbach et al., 2021)	Structure of RNA polymerase II pre-initiation complex at 2.9 A˚defines initial DNA opening	Structures show Tfb3 RING-Pol II stalk-TFIIE Tfa1 interface in higher resolution.	*S. cerevisiae*	7O4K (contracted TFIIH within PIC), 7O4L (expanded TFIIH within the PIC)	2021	9-138 (full 321)
(Aibara et al., 2021)	Structures of mammalian RNA polymerase II pre-initiation complexes	Structures show RING finger domain of MAT1 is displaced after CC-to-OC transition, therefore TFIIH detaches from cPIC	*H. sapiens* GTFs; *Sus scrofa domesticus* Pol II	7NVY (proximal CC), 7NW0 (OC)	2021	1-149 (full 309)

The N-terminal **RING finger domain** of Tfb3 (residues 1-69) comprises a conserved C3HC4 zinc motif, in which seven cysteines and one histidine are coordinated with two zinc ions (**Figure 7**). This RING finger is highly conserved from humans to yeast, especially the cysteines/histidine binding to zinc ions and a couple of hydrophobic residues stabilizing finger core (Gervais et al., 2001) (**Figure 7**). Specifically, the 1st, 2nd, 4th and 5th cysteines (C13, C16, C39, C42) of yeast Tfb3 RING finger bind zinc ion I (Zn-I), and the rest of the cysteines (C34, C39, C54) and the histidine (H36) bind zinc ion II (Zn-II). It seems that disruption of binding of Zn-II is more detrimental to the structure of the RING domain, whereas the structure around Zn-I may be stabilized by other interactions. Two mutations at cysteines that bind to Zn-II (C34S and C59Y) are lethal (Jona et al., 2002). Additionally, my *TFB3* screen (**Chapter 3**) detected mutations at cysteines binding to Zn-I causing transcription-related phenotypes, but no mutations at positions binding to Zn-II. That might be because mutations at those positions cause lethality, so they never get a chance to be detected.

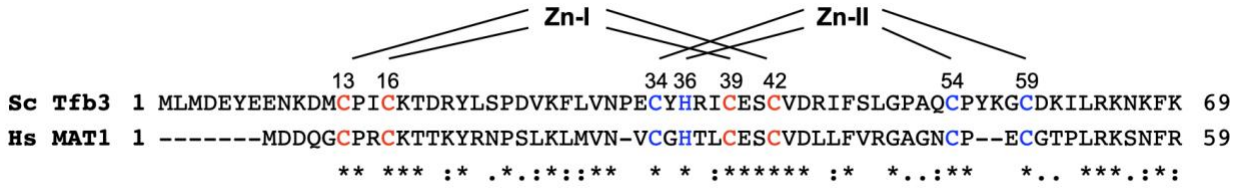


Figure 7 The RING finger domain of Tfb3 and its binding zinc ions

Alignment of amino acid sequences of RING finger domains of yeast Tfb3 and human MAT1, using local ClustalX. The residues binding to zinc ion I (Zn-I) and zinc ion II (Zn-II) are in red and blue, respectively. The conserved positions are indicated on the line below the alignment. The "*" indicates positions with a fully conserved residue. The ":" indicates positions that are strongly conserved, meaning belonging to a group exhibiting strong similarity. The "." indicates positions that are weakly conserved, meaning belonging to a group exhibiting weak similarity. The abbreviations and source of amino acid sequences used in alignment are as follows: Sc, *Saccharomyces cerevisiae*, SGD YDR460W; Hs, *Homo sapiens*, UniProtKB P51948.

The Tfb3 RING domain has been demonstrated to be involved in widely varying cellular functions: (1) linking TFIIH to Pol II stalk subcomplex Rpb4-Rpb7 and TFIIE Tfa1 within the PIC (Schilbach et al., 2017) (see next section), (2) facilitating Kin28 phosphorylation (Jona et al., 2002), (3) contributing to NER (Feaver et al., 2000), and (4) controlling activation of ubiquitin ligase (E3) cullins Cul3 and Rtt101, which activates E3 complexes assembly and ubiquitin transfer to the substrate (Rabut et al., 2011). Supporting these involved activities, the Tfb3 RING finger domain has been observed to cross-link or interact with (1) TFIIH core subunits Ssl2 (Schilbach et al., 2017), Rad3 (Luo et al., 2015; Robinson et al., 2016), Tfb1 (Luo et al., 2015), (2) TFIIH kinase module subunit Kin28 (Robinson et al., 2016), (3) Pol II stalk subunits Rpb4 (Schilbach et al., 2021) and Rpb7 (Schilbach et al., 2021; Schilbach et al., 2017), and (4) TFIIE subunit Tfa1 (Robinson et al., 2016; Schilbach et al., 2021; Schilbach et al., 2017).

Many crosslinks have been detected between the central linker region (including partial ARCH anchor domain and α-helical domain) and the C-terminal hydrophobic domain of Tfb3, therefore a "**Latch region**" (residues 120-314) was proposed as a structural motif (Luo et al., 2015; Robinson et al., 2016) (**Figure 6**). The Latch region was proposed to support a linkage between TFIIH kinase and core modules via interactions with core subunits Ssl1 (Luo et al., 2015), Ssl2 (Luo et al., 2015), and Rad3 (Luo et al., 2015; Robinson et al., 2016), and with kinase subunits Kin28 (Luo et al., 2015; van Eeuwen et al., 2021) and Ccl1 (Luo et al., 2015; Robinson et al., 2016; van Eeuwen et al., 2021). In addition, the Latch region has been suggested to have important roles in Pol II CTD, with detected crosslinking to Mediator subunits Med8 and Med11 and Pol II subunit Rpb4 (Robinson et al., 2016).

Several studies show that Tfb3 is involved in two activities of transcription initiation: promoter scanning and phosphorylation of Pol II CTD. Each of these activities will be discussed below.

1.3.3 Tfb3 bridges between TFIIH, Pol II, and TFIIE in the PIC

A critically important role of Tfb3 is its bridging of TFIIH, Pol II, and TFIIE within the PIC, which is completed by three parts of interactions (**Figure 8**): the Tfb3-Pol II Rpb4-Rpb7 interaction, the Tfb3-TFIIE Tfa1 interaction, and the Tfb3-TFIIH Rad3 interaction. First, Tfb3 bridges TFIIH and Pol II via its RING finger domain interacting with Pol II Rpb4-Rpb7 stalk subcomplex. This Tfb3-Rpb4-Rpb7 interaction mainly comprises a charged interface including three potential salt bridges: Tfb3 R64-Rpb7 D166 (**Figure 9A**), Tfb3 K65-Rpb7 E165 (**Figure 9A**), and Tfb3 K67-Rpb4 D189 (**Figure 9B)**. The first, Tfb3 R64-Rpb7 D166, is the most intriguing and provocative interaction. The distance of this salt bridge is 2.1-3.1 Å in two PIC structures (Schilbach et al., 2021; Schilbach et al., 2017). Substitutions on both sides of this salt bridge, *tfb3* R64K (unpublished, from *TFB3* screen in **Chapter 3**) and *rpb7* D166G (Braberg et al., 2013), showed MPA-sensitive phenotypes, which correlate with TSS upstream shifting and have been further proposed to indicate "Loss-of-Function" class of *tfb3* and *rpb7* mutants (see **Chapter 3**). The TSS selection effects from both sides of this bridge suggest the association between Pol II and TFIIH is important for initiation by promoter scanning. The second and third potential salt bridges need to be directly examined because they have been observed varying distances in different structures (Schilbach et al., 2021; Schilbach et al., 2017). In addition, another MPA-sensitive mutant (*tfb3* N66D) was observed in this interface (unpublished, from *TFB3* screen in **Chapter 3**), which further supports the importance of this interface. Second, the Tfb3 bridges

TFIIH and TFIIE via its interaction with TFIIE subunit Tfa1. This Tfb3-Tfa1 interaction is mainly between the Tfb3 RING domain and a hydrophobic pocket formed by Tfa1 E-linker helices together with a part of the Rpb7 OB domain. Interestingly, the involvement of this interface in TSS selection is confirmed by our screens for *TFB3* (see **Chapter 3**) and *TFA1* (unpublished data from the Kaplan lab). On both sides of Tfb3 and Tfa1, a couple of strong MPA-sensitive mutants (proposed LOF mutants) obtained from screens are within this interface. These mutations likely interfere with interactions and confer phenotypes predictive of upstream TSS shifting. Third, the Tfb3 tethers the TFIIH core subunits to the rest of the PIC via the interaction of its "ARCH anchor" (residues 70-146) with the TFIIH core subunit Rad3 ARCH domain (residues 249-441) (Robinson et al., 2016; Schilbach et al., 2021; Schilbach et al., 2017).

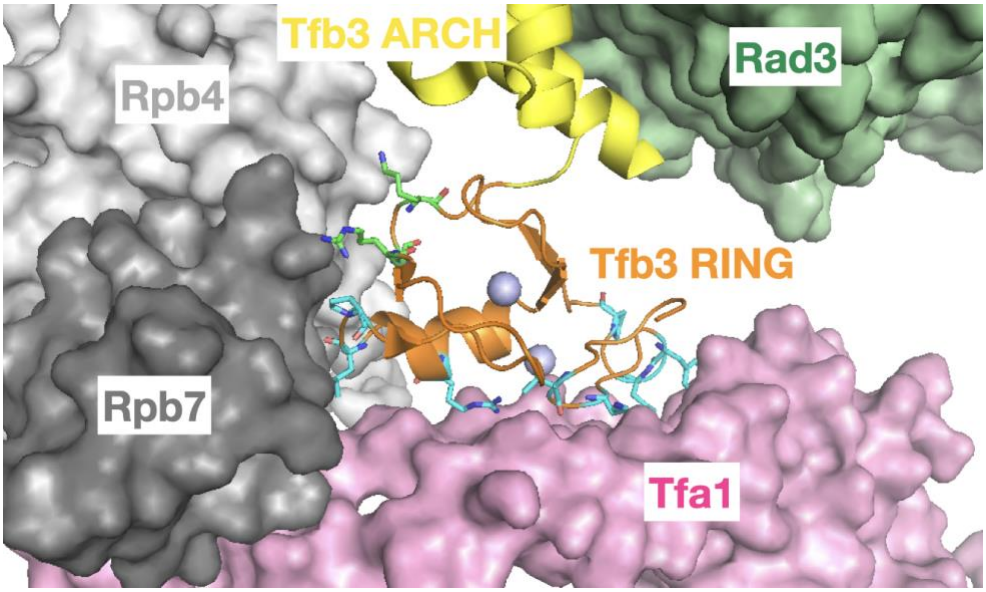


Figure 8 Tfb3 bridges between TFIIH, Pol II, and TFIIE in the PIC

The interactions of Tfb3 with the Pol II Rpb4-Rpb7 stalk subcomplex, TFIIE subunit Tfa1, and the TFIIH core subunit Rad3 (PDB: 7O4L). The RING finger and the ARCH anchor domains of Tfb3 are depicted as cartoons, and other factors are shown as surfaces. The proposed interface residues of the Tfb3 RING finger domain with Rpb4-Rpb7 stalk and Tfa1 are shown as sticks (Schilbach et al., 2021). Additionally, interface residues of Tfb3 involved in the charged interface with Rpb4-Rpb7 stalk and in a hydrophobic pocket of Tfa1 and Rpb7 are colored in green and cyan, respectively. The two zinc atoms associated with the RING finger domain are shown in light blue.

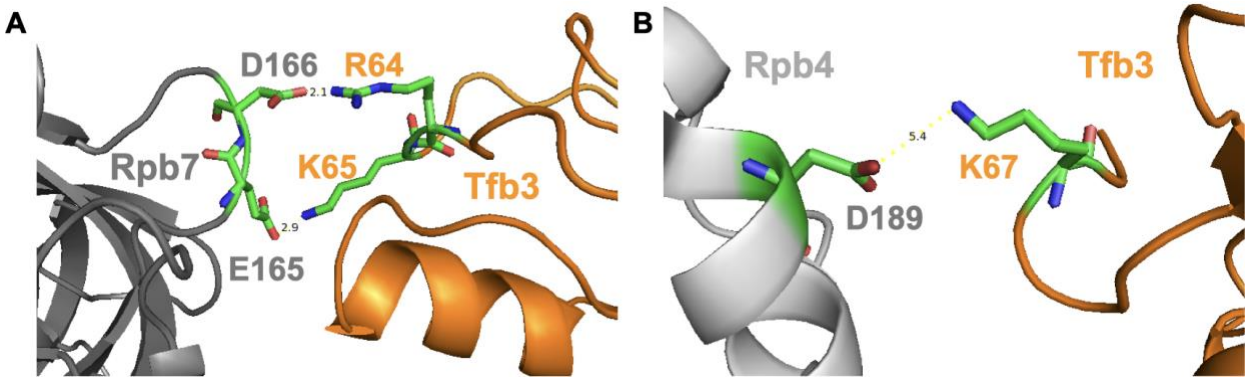


Figure 9 Three proposed salt bridges between the Tfb3 RING finger domain and Rpb4-Rpb7 stalk

The RING finger domain of Tfb3 interacts with the Pol II Rpb4-Rpb7 stalk via a charged interface, including **(A)** two proposed salt bridges between the RING finger and Rpb7 (PDB: 5OQJ) and **(B)** one proposed salt bridge between the RING finger and Rpb4 (PDB: 7O4L). The residues involved in potential salt bridges are shown as sticks.

How might we understand these interactions within the PIC and their relation to potentially different initiation models? A recent structure from a mammalian system indicates that the different strengths of these interactions in humans and yeast may be the reason for different types of initiation machinery (Aibara et al., 2021). In the mammalian PIC, DNA opening has been shown to initiate at approximately 30 bp downstream of the TATA-box. After the mammalian PIC transited from the closed promoter complex to the open promoter complex state, the MAT1 (Tfb3 in humans) has been shown to be detached from Rpb4-Rpb7 and TFIIE, which was proposed to be associated with TFIIH detachment from the PIC. The detachment of TFIIH has been predicted to prevent further downstream promoter scanning, therefore a TSS at 30-31 bp downstream of the TATA-box is typically used in humans. The interactions between MAT1/Tfb3 and Rpb4-Rpb7 and TFIIE are remarkably conserved in human and yeast PICs, except for some minor differences such as MAT1/Tfb3 interacts with stalk subunit Rpb7 only in the mammalian PIC but with both stalk subunits Rpb7 and Rpb4 in the yeast PIC (Aibara et al., 2021; Schilbach et al., 2021; Schilbach et al., 2017). This suggested similar but not identical functions of Tfb3 and MAT1 during initiation. The stronger contact between the Tfb3 and the Pol II stalk in the yeast PIC might make the detachment of TFIIH from the rest of the PIC occur later than what happens in humans, allowing downstream DNA scanning. Excitedly, my *TFB3* screen results in **Chapter 3** support this proposed model. All *tfb3* substitutions within either Tfb3-Pol II stalk or Tfb3-TFIIE interface that were predicted to disrupt linkages of TFIIH to the rest of the PIC showed MPA-sensitive phenotypes, with upstream shift effect on TSS confirmed for some alleles. This is consistent with the proposed model that easier dissociation of TFIIH from the rest of the PIC, caused by weaker contacts of Tfb3 with Pol II stack or TFIIE, leads to a shorter scanning distance. This proposed model is also in accordance with two previous observations related to TFIIK (including Tfb3,

Kin28, and Ccl1) omission. First, the lack of TFIIK shifts TSS usage upstream to the location that is for higher eukaryotes *in vitro* (Murakami et al., 2015), suggesting the requirement of TFIIK subunits for downstream TSS usage, though now Tfb3 is further proposed to be the regulator. Second, recent single molecule studies showed that there was no difference in DNA scrunching when TFIIK is absent in the PIC (Fazal et al., 2015), suggesting the omission of TFIIK did not affect the DNA translocation ability of TFIIH. Instead, it might be the coupling of TFIIH translocation ability to PIC scanning process that was affected by the loss of interactions via TFIIK (Tfb3). Taken together, a comprehensive study of Tfb3 function in yeast initiation and TSS selection will provide important sights into mechanism of initiation in yeast and reasons for different initiation models.

1.3.4 Tfb3 tethers the Kin28-Ccl1 kinase-cyclin pair to the PIC and facilitates Pol II CTD phosphorylation

At the end of transcription initiation, the Pol II subunit Rpb1 C-Terminal Domain (CTD) becomes phosphorylated by TFIIH kinase subunit Kin28, and subsequently Pol II escapes from the PIC and transits to elongation. This phosphorylation starts at Ser5 and can be enhanced by the presence of Mediator. To complete this CTD phosphorylation and therefore productive transcription, Kin28 needs to be phosphorylated and therefore be activated. In addition, the Kin28 active site needs to be placed in proximity to its substrate (Rpb1 CTD), which is likely achieved by a tunnel formed by TFIIK and Mediator (Abdella et al., 2021; Chen et al., 2021; Plaschka et al., 2015; Robinson et al., 2016; Schilbach et al., 2017; van Eeuwen et al., 2021).

Tfb3 has been suggested to contribute to Pol II Rpb1 CTD phosphorylation through multiple mechanisms. First, Tfb3 tethers the Kin28-Ccl1 kinase-cyclin pair to Pol II and Mediator

via interaction of the Tfb3 RING finger domain with the PIC, as described in the previous section, and the interaction of the Tfb3 C-terminal hydrophobic domain with both Kin28 and Ccl1 (Jona et al., 2002; Schilbach et al., 2021; Schilbach et al., 2017; van Eeuwen et al., 2021). Second, Tfb3 promotes phosphorylation and activation of Kin28, which is required for Pol II CTD phosphorylation, by linking Kin28 to the TFIIH core. The phosphorylation of Kin28 by Cak1 (CDK-activating kinase 1) requires the presence of Tfb3 and Ccl1 (Espinoza et al., 1998). A mutation within the Tfb3 RING finger resulting in a reduced level of Tfb3 reduced Kin28 protein association with TFIIH and reduced Kin28 phosphorylation, except on the small portion of Kin28 that was still associated with TFIIH (Jona et al., 2002). These results suggested that connecting Kin28 to TFIIH core by Tfb3 is required for efficient phosphorylation of Kin28. Third, the C-terminal region of Tfb3 interacts with Kin28 and Ccl1 to stabilize the Kin28 activation loop (known as T-loop) in its catalytically active conformation as well as locate the Kin28 T-loop to the edge of the catalytic cleft, therefore increasing the chance for CTD to access the T-loop active site (van Eeuwen et al., 2021).

1.4 Overview of dissertation

Most promoters in eukaryotes utilize multiple TSSs. As the first step of transcription, initiation determines where and how efficiently transcription initiates and therefore is a critical point of control in gene expression. Pol II initiation in yeast proceeds by a proposed promoter scanning mechanism, where the PIC, comprised of Pol II and GTFs, assembles upstream of the initiation region and then scans downstream to select appropriate TSSs to initiate transcription. What GTF(s) regulate the scanning process, how scanning is affected by or interacts with different

promoter architectural features (such as DNA sequence, core promoter-TSS distance), and how these architectural features interact with initiation factor activity (such as Pol II, GTFs), are important open questions in the field of transcription. To determine Pol II TSS sequence specificity and how Pol II activity alters it, I developed a massively parallel report assay (Pol II MASTER) and have measured initiation efficiency for ~1 million unique TSS sequences in WT and Pol II catalytic mutants (**Chapter 2**). To dissect the role of the TFIIH subunit Tfb3 and how it interfaces with other initiation factors (TFIIH subunit Ssl2 and PIC cofactor Sub1) to determine TSS selection, I screened for *tfb3* mutants conferring transcriptional defects, used genetic tools to examine their effects on TSS usage, and investigated their genetic relationships with *ssl2* and *sub1*Δ alleles (**Chapter 3**). To examine how other promoter architectural features determine Pol II initiation, including UAS identity, core promoter-TSS distance, and sequence composition within the scanning region, I designed and am constructing "architecture" libraries that apply developed Pol II MASTER analysis to other Pol II initiation regulatory elements (**Appendix A**).

2.0 Quantitative analysis of transcription start site selection in *Saccharomyces cerevisiae* determines contributions of DNA sequence and RNA Polymerase II activity

Transcription Start Site (TSS) selection is a key step in gene expression and occurs at many sites over a wide range of efficiencies. Here, we develop a massively parallel reporter assay to quantitatively dissect contributions of promoter sequence and RNA Polymerase II (Pol II) activity to TSS selection by "promoter scanning" in *Saccharomyces cerevisiae* (Pol II MAssively Systematic Transcript End Readout, "Pol II MASTER"). Using Pol II MASTER, we measure the efficiency of Pol II initiation at 1,000,000 individual TSS sequences in a defined promoter context. Pol II MASTER not only recapitulates known critical qualities of *S. cerevisiae* TSS -8, -1, and +1 positions but also demonstrates that surrounding sequences modulate initiation efficiency over a wide range. We discover functional interactions between neighboring sequence positions, indicating that adjacent positions likely function together. The results enable the development of a predictive model for initiation efficiency by promoter scanning at genomic promoters. We show that genetic perturbation of Pol II catalytic activity alters initiation efficiency mostly independently of TSS sequence, but selectively modulates preference for the initiating nucleotide. The results establish Pol II MASTER as a method for quantitative dissection of transcription initiation in eukaryotes.

2.1 Introduction

In transcription initiation, RNA Polymerase II (Pol II) binds promoter DNA through interactions with core promoter elements, unwinds a turn of promoter DNA forming a Pol II-promoter open complex containing a single-stranded "transcription bubble", and selects a promoter position within a region competent for initiation to serve as the Transcription Start Site (TSS). At the majority of Pol II promoters in eukaryotes, TSS selection occurs at multiple positions (Carninci et al., 2005; Chen et al., 2013; Chia et al., 2021; Consortium et al., 2014; Hoskins et al., 2011; Nepal et al., 2013; Park et al., 2014; Pelechano et al., 2013; Yamashita et al., 2011; Zhang and Dietrich, 2005; Zheng et al., 2011). Thus, the overall rate of gene expression at the majority of Pol II promoters is determined by the efficiency of initiation from several distinct TSS positions. In addition, studies have suggested that alternative TSS selection can lead to differences in mRNA features, translation activity, and subsequent protein levels and functions (Cheng et al., 2018; Pelechano et al., 2013; Rojas-Duran and Gilbert, 2012), and therefore is widespread in different cell types (Consortium et al., 2014), developmental processes (Batut et al., 2013; Cheng et al., 2018; Chia et al., 2021; Zhang et al., 2017), growth conditions (Lu and Lin, 2019), responses to environmental changes, and cancers (Boyd et al., 2018; Demircioglu et al., 2019; Thorsen et al., 2011).

Pol II initiation in yeast proceeds by a promoter scanning mechanism, where the Pol II Pre-initiation Complex (PIC), comprising Pol II and initiation factors, assembles upstream of the initiation region and then scans downstream to select a TSS position (Fazal et al., 2015; Hampsey, 1998; Kaplan et al., 2012; Kuehner and Brow, 2006; Lis, 1993; Miller and Hahn, 2006; Qiu et al., 2020; Zhao et al., 2021). The efficiency of initiation at a given position within a region allowing initiation depends on its location relative to the core promoter region, with bases scanned from

upstream to downstream from the core promoter, and on DNA sequence, with the template base specifying the TSS position and position immediately upstream of the TSS (positions +1 and -1, respectively) making the largest contributions. In particular, there is a strong preference for an R:Y base pair at position +1 and Y:R base pair at position -1 (reflected as a $Y_{-1}R_{+1}$ "initiator" sequence on the coding strand; Y = pYrimidine; R = puRine). Furthermore, this is a near universal preference for transcription initiation by multi-subunit and single-subunit RNA polymerases (RNAPs) in all domains of life (Bucher, 1990; Carninci et al., 2006; Chen et al., 2013; Corden et al., 1980; Cortes et al., 2013; Furter-Graves and Hall, 1990; Hahn S, 1985; Hashimoto et al., 2004; Healy et al., 1987; Kim et al., 2012; Kuehner and Brow, 2006; Lu and Lin, 2019; Malabat et al., 2015; McNeil and Smith, 1985; Policastro et al., 2020; Qiu et al., 2020; Smale and Baltimore, 1989; Suzuki et al., 2001; Vvedenskaya et al., 2015b; Zhang and Dietrich, 2005; Zheng et al., 2011). Structural work indicates that preference for a $Y_{-1}R_{+1}$ initiator sequence occurs, at least in part, by stacking of a purine NTP bound to template strand position +1 and a purine base at template strand position -1 (Basu et al., 2014; Gleghorn et al., 2011).

It has long been recognized that DNA sequences at positions other than -1/+1 also contribute to the efficiency of initiation by Pol II. However, the sequence preferences at these positions are not well understood, might extend to higher-order interactions between positions, and likely include species-specific determinants. For example, in *S. cerevisiae*, genome-wide transcriptomic data have revealed a preference for an A:T base pair at position -8, reflected as an A on the coding strand and a T on the template strand (Lu and Lin, 2019; Malabat et al., 2015; Policastro et al., 2020; Qiu et al., 2020; Zhang and Dietrich, 2005). Furthermore, promoters with high expression and with initiation focused primarily in a single, efficient TSS tend to show additional A enrichment at positions -7 to -3 on the coding strand (T on the template strand) (Lu

and Lin, 2019; Lubliner et al., 2013; Maicas and Friesen, 1990; Qiu et al., 2020; Zhang and Dietrich, 2005). Several factors confound efforts to directly measure the contribution of DNA sequence to the efficiency of initiation by Pol II, including promoter chromatin context and TSS position within a promoter. These attributes together might be considered a promoter's architecture. While promoter architectural factors are made especially apparent in yeast – where initiation occurs by promoter scanning – and TSSs are examined in a polar fashion from upstream to downstream, they likely contribute to the efficiency of TSS usage by Pol II in most, if not all, eukaryotes.

An elegant and important study of promoter scanning from Kuehner and Brow established that TSS usage is determined by TSS priority during the scanning process (Kuehner and Brow, 2006). As noted above, scanning proceeds directionally from upstream sequences to those downstream. Therefore, sequences are examined by the transcription machinery in the order in which they are scanned and upstream sequences will have priority over downstream ones, regardless of innate TSS strength. To enable comparison of TSSs with different usages in different positions, Kuehner and Brow introduced a concept of "TSS efficiency," which accounts for how much Pol II reaches a particular TSS in order to determine innate TSS strength (see **Figure 10**).

Imbalanced promoter sequence distributions imposed by evolutionary processes also limit our ability to determine sequence-activity relationships for initiation. For example, it has been observed that yeast promoters have an uneven distribution of bases across promoters and this is most obvious in enrichment for T on the coding strand upstream of the median TSS position and A on the coding strand downstream of the median TSS position at highly expressed promoters, and a paucity of G/C content in general (Dujon, 1996; Lu and Lin, 2020; Lubliner et al., 2013; Maicas and Friesen, 1990; Qiu et al., 2020). We previously observed that Pol II activity mutants

conferred changes to TSS selection, including apparent alterations in usage levels for subsets of TSS motifs. Pol II catalytically hyperactive mutants (termed GOF for "Gain-of-Function") and hypoactive mutants (termed LOF for "Loss-of-Function") showed overall decreased and increased usage for TSSs with an A at position -8 relative to the TSS (-8A TSS), respectively (Qiu et al., 2020). However, the biased distribution of bases in yeast promoters leads to a biased distribution of sequence motifs, where preferred TSS motifs show enrichment downstream of the median TSS position and less-preferred motifs show enrichment upstream of the median TSS position (Qiu et al., 2020). Because Pol II GOF and LOF mutants were observed to shift TSS usage distributions upstream or downstream, respectively, we could not determine whether the apparent altered TSS motif preferences were a direct or an indirect consequence of Pol II catalytic mutant effects on TSS selection. Moreover, other properties such as biochemistry of scanning processivity (Fazal et al., 2015; Lis, 1993; Zhao et al., 2021), promoter identity (Blazeck et al., 2012; Dhillon et al., 2020; Lubliner et al., 2015), or promoter chromatin could also contribute to initiation output. Together, all of these factors would make contribution of primary DNA sequence more difficult to ascertain.

In order to remove contextual differences among promoters we have developed a system to dissect determinants of initiation efficiency within a defined, controlled promoter context. Here, we develop "Pol II MASTER" based on bacterial MASTER (MAssively Systematic Transcript End Readout) (Hochschild, 2015; Vvedenskaya et al., 2018; Vvedenskaya et al., 2016; Vvedenskaya et al., 2015b; Winkelman et al., 2016), which allows determination of initiation at base pair resolution and attribution of RNA transcripts to nucleic acid barcoded promoter variants in a massively parallel fashion. We apply Pol II MASTER to initiation by promoter scanning to investigate the initiation efficiency of ~80,000 promoter variants within an appropriate TSS region

in Pol II WT and catalytic mutants. We show that this system enables determination of the interface between initiation factor activity, promoter sequence, and promoter output. We recapitulate the large impact of known base positions relative to the TSS on initiation efficiency (-8, -1/+1) while revealing the wide range of effects other positions have on initiation (-11 to -9, -7 to -2, and +2 to +4). We identify a distinct hierarchy in $Y_{-1}R_{+1}$ preferences, and detect interactions between bases at neighboring positions, suggesting potential mechanistic coupling between positions. We find that Pol II mutant classes increase or decrease initiation efficiencies for all possible sequences, consistent with predictions that the primary effects of altered TSS selection (directional shifts in TSS distributions) are driven by initiation efficiency changes across all sequences and not on TSS sequence selection per se. Our results further show that Pol II activity level does contribute to selective efficiency of initiation at sequences +1A vs +1G. Our findings demonstrate that Pol II MASTER provides a platform for quantitative analysis of how initiation factor activity together with promoter sequence contributes to Pol II transcription initiation *in vivo*.

2.2 Results

2.2.1 A high-throughput system for studying TSS sequence effects on Pol II initiation output

TSSs are identified at yeast promoters by scanning from upstream near the core-promoter to downstream (**Figure 10A**). While promoters are melted starting around +20 from the TATA box (if present), there is a distance restriction that reduces usage of TSSs until they are within a region ~40-140 nt downstream from the core promoter (**Figure 10B**). Downstream DNA is

pumped toward the Pol II active site by General Transcription Factor (GTF) TFIIH. As TFIIH remains attached to the PIC, this causes scrunching of DNA within the PIC. One model for this scrunching suggests a large transcription bubble with excess DNA being looped out of the PIC. As the scanning process is controlled biochemically by DNA translocation driven by TFIIH, this means that some TSSs may simply be too far downstream to have a high probability of usage (**Figure 10A**, "unreachable TSS"). Once initiation happens at any site within this window of opportunity for initiation (**Figure 10A, B**), "Pol II flux" – the amount of Pol II progressing to TSSs downstream – is necessarily reduced. Differences between promoters in TSS positioning relative to core promoter proximity restriction, TFIIH processivity, and differences in Pol II flux due to usage at other TSSs mean that DNA sequence effects on initiation are difficult to distinguish. In order to specifically dissect how TSS sequence controls initiation efficiency and how sequence interacts with Pol II catalytic activity in a controlled context, we have established a massively parallel promoter variant assay "Pol II MASTER" where we embed almost all possible sequences within a 9 bp randomized TSS region (**Figure 10C**) constructed on plasmids and introduced into yeast strains with wild type (WT) or mutated Pol II. The sequence libraries constructed are illustrated in **Figure 10C** and are referred to based on their base compositions relative to the coding strand unless otherwise specifically noted. The libraries are referred to specifically based on the identities of bases at positions -8, -1, and +1. The "AYR" library has composition $A_{-8}NNNNNNY_{-1}R_{+1}$ (N = A, C, G, or T, Y = C or T, R = A or G) relative to the coding strand, with "BYR" having composition $B_{-8}NNNNNNY_{-1}R_{+1}$ (B = C, G, or T), etc.

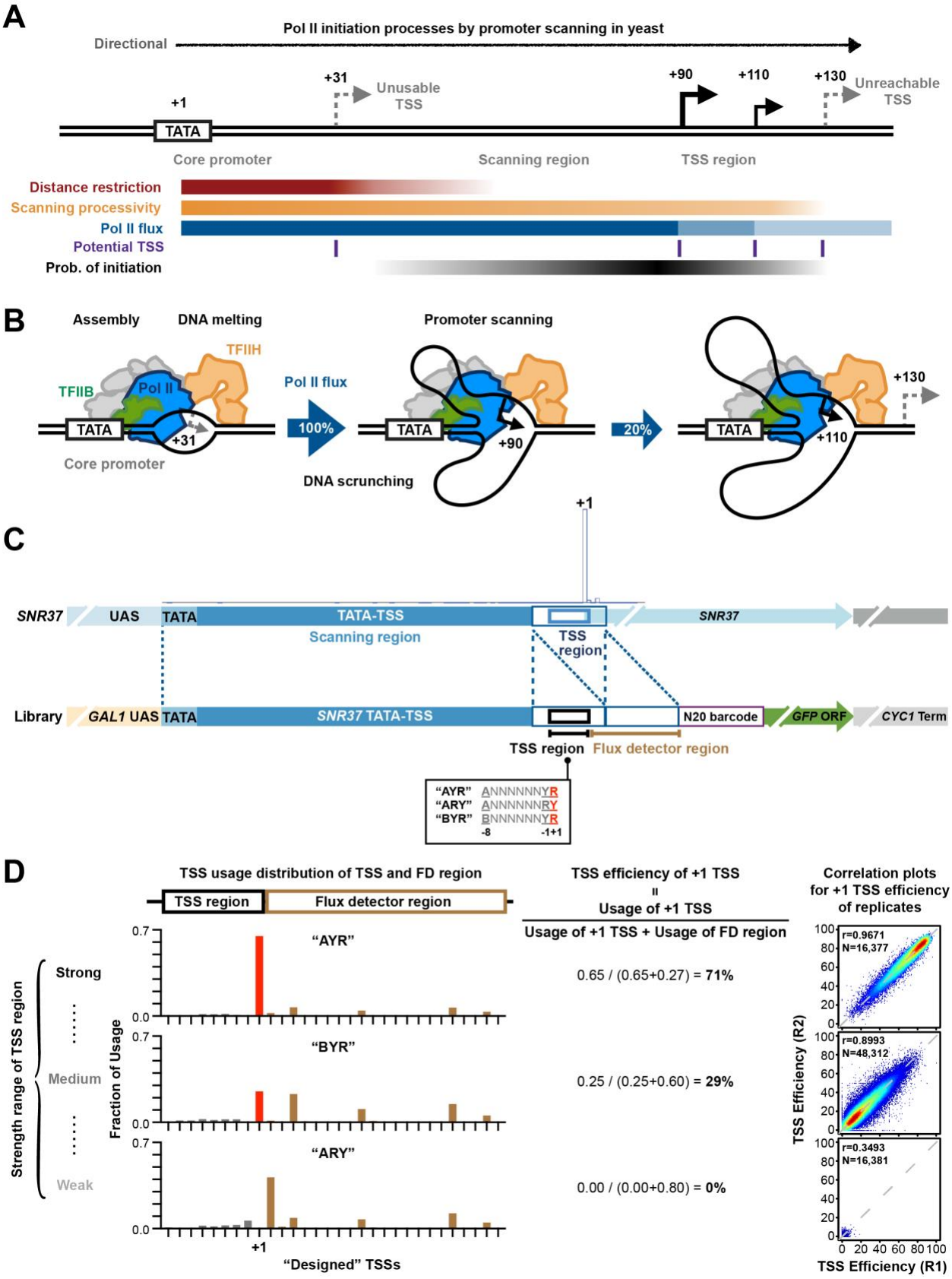


Figure 10 A high-throughput system for studying transcription TSS selection.

(A and B) Pol II initiation in yeast proceeds by a promoter scanning mechanism. Yeast Pol II initiation usually occurs at multiple TSSs ranging from 40 to 120 bp downstream of a core promoter region comprising the PIC assembly position and typically a TATA-box for TATA-containing promoters. After the PIC assembles upstream (core promoter), scanning will proceed over a scanning region toward positions where TSS selection occurs (TSS region). In addition to sequence control of initiation through suitable TSS sequences, the probability of initiation across promoter positions is also controlled by multiple architectural features shown in **(A)**. These include the inhibition of initiation near a core promoter that diminishes as scanning proceeds ("distance restriction"), biochemical restrictions on how far scanning can proceed that increase as scanning proceeds ("scanning processivity"), and "Pol II flux", which represents the decrease in amount of scanning Pol II as scanning proceeds due to conversion of scanning Pol II to transcribing Pol II upon initiation. **(C)** Construction of promoter libraries that control TSS sequence context. The top panel shows a schematic of the *SNR37* promoter and its TSS usage distribution based on TSS-seq (Qiu et al., 2020). *SNR37* has a focused and highly efficient TSS region. The bottom panel shows a schematic of the Pol II MASTER libraries used in this study. A duplication of the *SNR37* TSS region was inserted before the native TSS region, and the -8 to +1 positions relative to native *SNR37* +1 TSS (black box) were replaced by a 9 nt randomized region to provide almost all possible sequences. The second *SNR37* TSS region functions as a "Flux Detector" (FD) to capture Pol II flux that fails to initiate within the randomized region and allow determination of initiation efficiency within the randomized region. A barcode region (purple box) allows RNA products to be assigned to respective promoter variants. The *GAL1* UAS allows control of library expression, and the *GFP* ORF and *CYC1* terminator support termination and stabilize RNA products. **(D)** TSS usage distributions at TSS and FD regions for different promoter variant "AYR", "BYR", and "ARY" libraries are shown on the left. TSS usages from designed +1 TSS and positions upstream are in red and grey, respectively. TSS usage from the FD region is in brown. The definition of "TSS efficiency" and overall TSS efficiency calculations for the aggregate +1 TSS of all in different libraries are shown in the middle. Example correlation plots of TSS efficiency calculations for +1 TSSs from individual promoter variants in Pol II MASTER libraries between representative biological replicates are shown on the right. Pearson r and the number of compared variants are shown.

This 9 bp randomized TSS region was inserted into a controlled promoter context containing specific functionalities (**Figure 10C**). First, the *GAL1* UAS was utilized to allow

expression control of libraries. Second, the TATA-box to TSS region of the *SNR37* gene was used as a "scanning region" to direct initiation within the adjacent randomized TSS region. This is because almost no RNA 5′ ends are observable from this scanning region within its normal promoter context. Third, the native, highly-efficient *SNR37* TSS region was inserted downstream of the randomized TSS region as a "Flux Detector" (FD). Here, we employ the approach of Kuehner and Brow where a highly efficient initiation region placed downstream of a TSS region may capture any polymerases that happen to scan past the randomized TSS region. Therefore, TSS efficiency, as defined in (Kuehner and Brow, 2006), can be measured for any TSS of interest as Pol II scanning past the TSS should effectively be specified to initiate (**Figure 10C**). Without this FD region, an absence of downstream initiation would render any upstream TSS as apparently highly efficient as there would be no point of reference. By measuring TSS efficiency using our FD, we can compare TSSs at different upstream positions within promoters and/or across libraries. Fourth, a 24 bp DNA barcode containing 20 positions of randomized bases and 4 interspersed fixed bases (to exclude low-complexity sequences) allows RNA products to be assigned to respective randomized TSS DNA templates. An RNA barcode is critical as bases upstream of the TSS will not be present in the transcribed RNA but are critical for specifying TSS efficiency. Fifth, the *GFP* coding region and *CYC1* terminator have been added to support termination and stabilize RNA products. Libraries were constructed by PCR sewing followed by cloning into plasmid backbone. After amplification in *E. coli*, plasmid libraries were transformed into Pol II WT and mutant yeast strains in triplicate. Library expression was induced by addition of galactose to the medium (4% final) for three hours. Both plasmid DNA and RNA products were extracted from harvested yeast cells and amplified for DNA-seq and TSS-seq (**Figure 11A**).

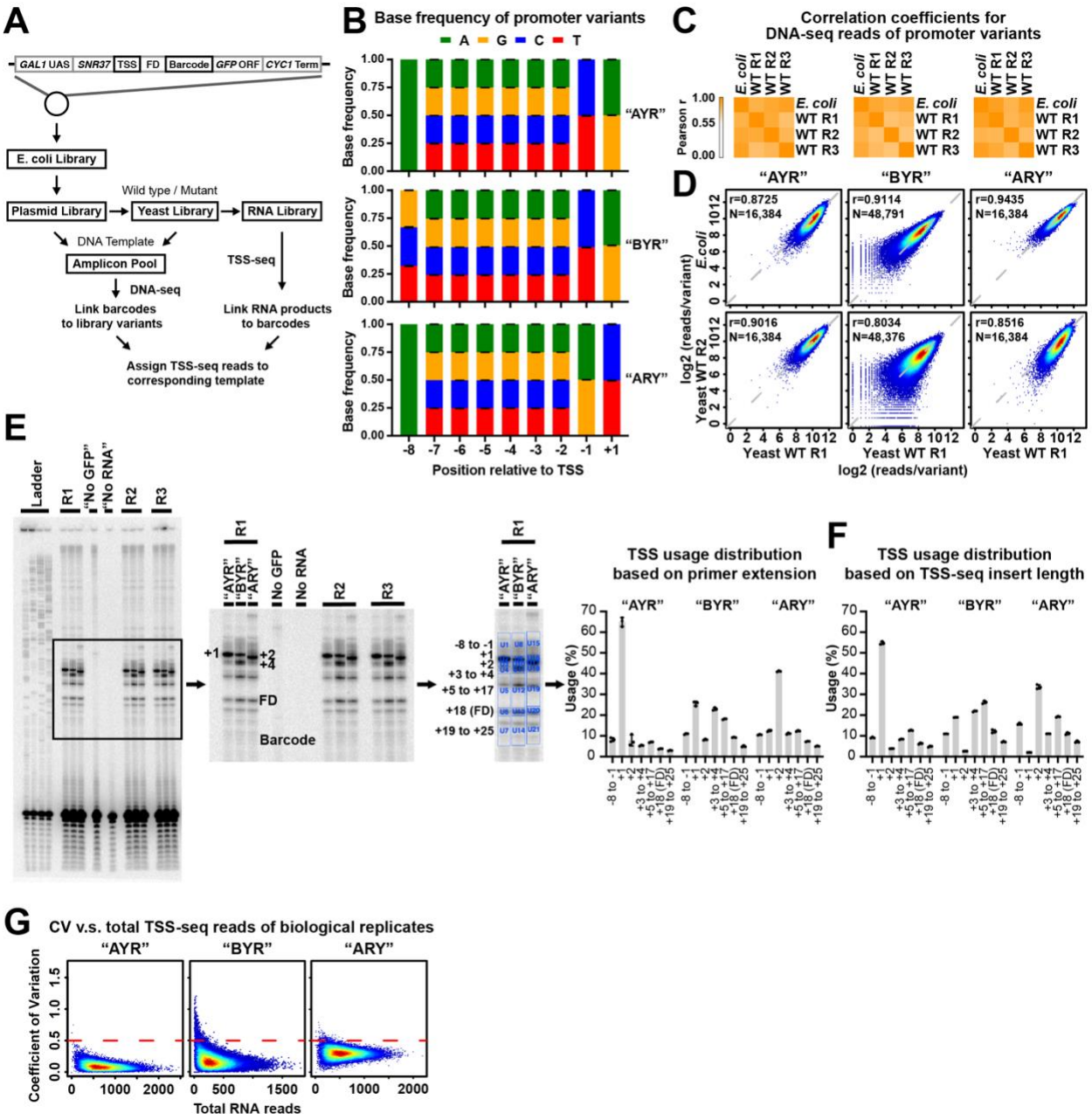


Figure 11 High level of reproducibility and coverage depth of library variants.

(A) Schematic representation of experimental approach. Promoter libraries with almost all possible sequences within a 9 nt randomized region within a promoter context designed for specific functionalities were constructed on plasmids. Libraries were of three types, designated "AYR", "BYR", and "ARY" based on compositions of their randomized regions. Plasmids were amplified in *E. coli* and transformed into yeast strains with wild type or mutated Pol II. DNA and RNA were extracted from yeast pellets and prepared for DNA-seq and TSS-seq. **(B)** Base frequencies at positions within the randomized region of promoter variants indicating unbiased coverage of randomized regions. Bars are mean

+/- standard deviation of mean of promoter variants in WT and four Pol II mutants. **(C)** Heatmap illustrating hierarchical clustering of Pearson correlation coefficients of reads per promoter variant for libraries amplified in *E. coli* and three biological replicates of these same libraries transformed into yeast. **(D)** Example correlation plots of DNA reads count of promoter variants for *E. coli* and yeast WT biological replicates. Pearson r and the number of compared variants are shown. **(E)** Bulk primer extension for RNA produced from promoter variant libraries and quantification for biological replicates transformed into WT yeast. "No GFP" control used an RNA sample without library transformed. "No RNA" control used a sample of nuclease-free water. Dots represent three biological replicates. Bars are mean +/- standard deviation of mean. **(F)** TSS usage distribution based on insert length of TSS-seq reads generated from transformed libraries. Dots represent three biological replicates. Bars are mean +/- standard deviation of mean. Compare similarity to distribution in **(E)**. Note that primer extension will blur usage into adjacent upstream position due to some level of non-templated addition of C to RNA 5′ ends. **(G)** Heat scatter plots of Coefficient of Variation (CV, y-axis) versus total RNA sequencing reads per promoter variant in each of three Pol II MASTER libraries. A cutoff of CV = 0.5 was used to filter out variants with higher variance.

Several measures indicate a high level of reproducibility and coverage depth of library variants (**Figure 10D, Figure 11**). Base coverage in the randomized region was highly even (**Figure 11B**). Correlation analysis for DNA-seq reads per promoter variant suggested that yeast transformation did not alter promoter variant distribution (**Figure 11C, D**). Bulk primer extension of libraries illustrated their average behavior and the amount of initiation derived from the randomized region (**Figure 11E**). As designed, only a very small fraction of initiation was generated from the barcode region or further downstream, validating that the flux detector captured scanning polymerases (**Figure 11E, F**). Aggregate distribution of reads in our three libraries shows that as TSSs decreased in efficiency from the most efficient library ("AYR") to least ("ARY"), reads shift from the designed +1 TSS to downstream positions (**Figure 10D**, left). The apparent shift to upstream position -1 in the ARY library is because purine at designated -1 position serves as the +1 for newly created TSSs. **Figure 10D** (middle) illustrates the aggregate TSS efficiency of

each library based on usage at the designated +1 TSS relative to usage at that position plus all downstream usage. As an example of high reproducibility, correlation analysis for efficiency of library "major" TSSs (designed +1 TSS of "AYR" and "BYR" libraries, +2 TSS of "ARY" library) demonstrated that biological replicates were reproducible (**Figure 10D**, right). Therefore, we summed reads from three biological replicates, keeping TSSs that contained at least five TSS-seq reads in each replicate and whose Coefficient of Variation (CV) of TSS-seq reads across replicates was less than 0.5 as a proxy for reproducible behavior (**Figure 11G**). As a result, 97% of possible TSS promoter variants were covered in each library on average (**Table 2**).

Table 2 Summary of libraries

Pol II	Library	Num. of promoter variants	Coverage
WT	AYR	16,377	99.96%
	BYR	47,556	96.75%
	ARY	16,224	99.02%
E1103G	AYR	16,352	99.80%
	BYR	43,947	89.41%
	ARY	16,179	98.75%
F1086S	AYR	15,886	96.96%
	BYR	47,403	96.44%
	ARY	15,669	95.64%
G1097D	AYR	16,080	98.14%
	BYR	46,119	93.83%
	ARY	16,152	98.58%
H1085Q	AYR	16,301	99.49%
	BYR	46,885	95.39%
	ARY	16,250	99.18%

2.2.2 Sequence-dependent control of TSS efficiency in *S. cerevisiae*

To ask how our libraries recapitulated known TSS efficiency measurements, we first examined core sequences in our library for the *SNR14* TSS and variants examined by Kuehner and Brow (Kuehner and Brow, 2006) for TSS efficiency (**Figure 12A**). Our randomized library contains the *SNR14* TSS sequence embedded in our *SNR37* context along with all single substitution variants of this sequence, including the subset previously examined in *SNR14.* We found that Pol II MASTER recapitulated the single base effects on TSS efficiency previously observed while also indicating that single base changes around a TSS can have large effects on TSS efficiency.

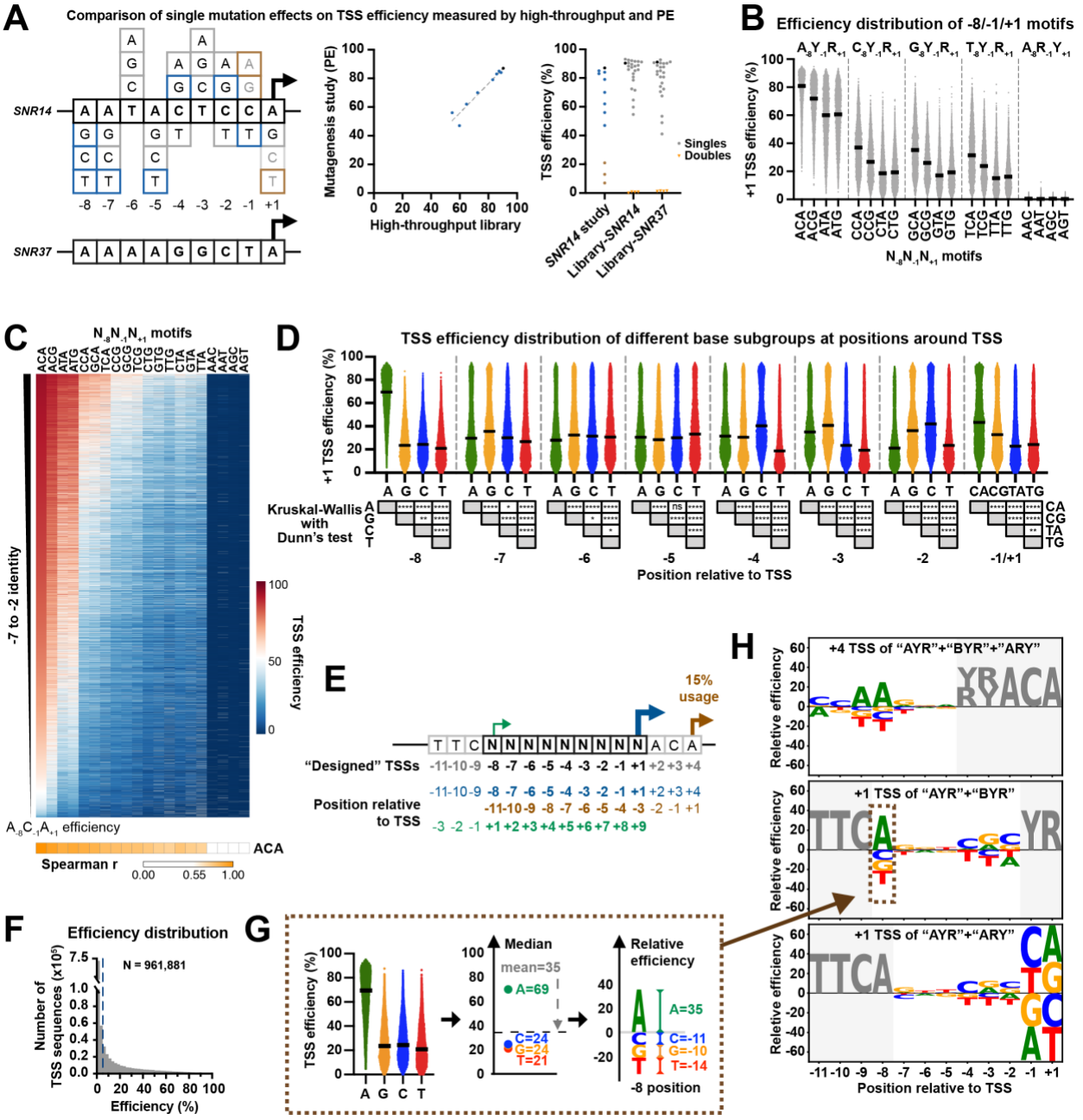


Figure 12 TSS sequence and its surrounding sequences modulate initiation efficiency over a wide range.

(A) Comparison of single mutation effects on TSS efficiency measured by high-throughput and primer extension. (Left) Sequences of *SNR14* and *SNR37* TSS regions (in black boxes, including positions between -8 to +1 relative to TSS) and all possible single substitutions of *SNR14* TSS region. Single substitutions included by both a prior *SNR14* mutagenesis study (Kuehner and Brow, 2006) and Pol II MASTER libraries are in blue while those lacking in our study are in brown. Additional substitutions analyzed here are in gray. (Middle) High correlation of TSS efficiency measured by Pol II MASTER and primer extension. Mutation classes are color coded as on left. (Right) Range of

effects of single base substitutions on TSS efficiency for *SNR14*- and *SNR37*-related sequences. Mutation classes are color coded as on left. Single substitutions absent from Pol II MASTER because of library design (R_{-1} and Y_{+1}) were expected to have super low TSS efficiencies. Double substitutions of *SNR14* and *SNR37* TSS region included in Pol II MASTER "ARY" library are shown as orange inverted triangles and show almost no efficiency. **(B)** Pol II initiation shows a strong preference for A_{-8} and $C_{-1}A_{+1}$ containing variants. All promoter variants were divided into 20 groups defined by bases at positions -8, -1, and +1 relative to the designed +1 TSS, and their +1 TSS efficiencies were plotted as spots. Lines represent median efficiencies of each group. **(C)** +1 TSS efficiency of all -7 to -2 sequences within each $N_{-8}N_{-1}N_{+1}$ motif in WT, rank ordered by efficiency of their $A_{-8}C_{-1}A_{+1}$ version, is shown as a heatmap. The x-axis is ordered based on median efficiency values for each $N_{-8}N_{-1}N_{+1}$ motif group, as shown in **B**. Statistical analyses by Spearman's rank correlation test between $A_{-8}C_{-1}A_{+1}$ group and all groups are shown beneath the heatmap. **(D)** Efficiency distributions of designed +1 TSSs grouped by base identities between -8 and +1 positions. Statistical analyses by Kruskal-Wallis with Dunn's multiple comparisons test for base preference at individual positions relative to +1 TSS are shown beneath data plots. Lines represent median values of subgroups. ****, $P \leq 0.0001$; ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$. **(E)** Definition of "designed" +1 TSS (designated as +1) and positions relative to this TSS (blue TSS arrow and sequence). TSS usage generated from upstream or downstream of "designed" +1 TSS (green and brown TSS arrows and sequences, respectively) allows us to study sequence preferences at positions -11 to -9 and +2 to +9 relative to the site of initiation. **(F)** Histogram showing the distribution of measured efficiencies for all designed -8 to +4 TSSs of all promoter variants from "AYR", "BYR" and "ARY" libraries in WT. Dashed lines mark the 5% efficiency cutoff used to designate a TSS as active. The total number of TSS sequences is shown. **(G)** Schematic illustrating how "relative efficiency" is calculated and visualized as sequence logo in **H**. At a particular position relative to a particular TSS, first, all variants were divided into four base subgroups defined by base at this position. Next, median values of each base group were extracted and centered based on the mean of all median values. The centered median values were defined as "relative efficiencies", representing preferences for bases at this particular position. Finally, relative efficiencies of bases were visualized as sequence logos. Positive and negative values indicate relatively preferred or less preferred bases. **(H)** Pol II initiation shows distinct sequence preference at positions around the TSS. The top panel was generated using datasets of designed +4 TSSs deriving from "AYR", "BYR" and "ARY" libraries. The middle panel was generated using datasets of designed +1 TSSs deriving from "AYR" and "BYR"

libraries. The bottom panel was generated using datasets of designed +1 TSSs deriving from "AYR" and "ARY" libraries. Positions that contain fixed or not completely randomized bases are shown in grey.

Examining all designated +1 TSS variants present in our library, we focused our analysis on positions -8, -1, and +1, which *in vivo* genome-wide data suggested are important determinants for TSS selection (**Figure 12B, C**). We first determined the TSS efficiencies of the designed +1 TSS for all promoter variants, dividing all variants into groups defined by bases at positions -8, -1, and +1 relative to TSS to examine the TSS efficiency distribution within each subgroup. Our results not only recapitulate the importance of these three positions in our defined promoter context but also demonstrate that surrounding positions have a considerable impact on TSS efficiency (note the wide range of efficiencies within each TSS group, **Figure 12B**). First, in our controlled context for TSSs at the designed +1 position, we found that $Y_{-1}R_{+1}$ was essential for initiation above a minimal background relative to $R_{-1}Y_{+1}$. Even in the presence of an A_{-8}, $R_{-1}Y_{+1}$ variants showed essentially no usage. Second, we quantified the very large effect of -8A on TSS efficiency (note that $A_{-8}Y_{-1}R_{+1}$ motifs were much higher in efficiency in aggregate than non-$A_{-8}Y_{-1}R_{+1}$ TSSs), demonstrating that -8A alters TSS efficiency. This result is in line with the observation that $A_{-8}C_{-1}A_{+1}$ motif-containing TSSs have the highest aggregate TSS usage from genomic promoters and appear to be the most efficient (Qiu et al., 2020). However, usage is a consequence of efficiency and promoter expression level, and genomic promoter efficiency analyses cannot account for other potential sequences or contributions. Third, among $Y_{-1}R_{+1}$ elements ($C_{-1}A_{+1}$, $C_{-1}G_{+1}$, $T_{-1}A_{+1}$, and $T_{-1}G_{+1}$) we found a clear hierarchy of efficiency that was not apparent from genomic promoter TSS usage or contexts, likely due to promoter sequence skew at these promoters. Previous genomic studies indicated $T_{-1}A_{+1}$ and $T_{-1}G_{+1}$ containing TSSs had higher aggregate usage than $C_{-1}G_{+1}$ containing TSSs (Lu and Lin, 2019; Qiu et al., 2020). We found that $C_{-1}G_{+1}$ containing TSSs were

more efficient than $T_{-1}A_{+1}$ and $T_{-1}G_{+1}$ containing TSSs, suggesting that lower aggregate usage of $C_{-1}G_{+1}$ containing TSSs at genomic promoters likely reflects A/T richness of the yeast genome (Dujon, 1996).

As noted above, a wide range of TSS efficiencies were observed within each -8/-1/+1 group, even the most efficient $A_{-8}C_{-1}A_{+1}$ group, suggesting a meaningful contribution of other positions to initiation efficiency. To ask whether the effects of individual -7 to -2 sequences have similar effects regardless of -8, -1, and +1 identity, we rank ordered all individual TSSs by the efficiency of their $A_{-8}C_{-1}A_{+1}$ version (**Figure 12C**). This rank ordering by $A_{-8}C_{-1}A_{+1}$ was predictive of efficiency ranks for -7 to -2 sequences with different bases at positions -8, -1, and +1. This observation indicates important contributions from positions beyond positions -8, -1, and +1 and that positions might function independently to determine TSS efficiency (examined below).

We set out to determine the contributions to Pol II TSS efficiency of individual bases at each position across our randomized region relative to the designed +1 TSS. To do so, we calculated the TSS efficiencies of examined TSS variants and divided them into subgroups based on bases at individual positions relative to the designed TSS +1 position (**Figure 12D**). Comparison of TSS efficiencies across base subgroups suggested significant individual base effects on TSS efficiency in aggregate at all examined positions. The usage of other positions in our promoter outside of the designed TSS +1 position (either within the randomized region or outside of it) allowed us to examine the contribution of bases in our randomized region on the efficiency of these other TSSs (**Figure 11E, F, Figure 12E**). Our combined libraries allowed us to analyze efficiencies of nearly a million TSS variants present within them (distribution of efficiencies is shown in **Figure 12F**). In order to visualize sequence preferences, we used the median initiation efficiency values of each base subgroup as indicators for preference, using

centered median values to calculate "relative efficiency", and illustrated this preference in a sequence logo (**Figure 12G**). Datasets of designed +1 TSSs deriving from "AYR", "BYR", and "ARY" libraries allowed us to nearly comprehensively study preference at positions -8 to +1 (**Figure 12H**, middle and bottom). Additionally, 10-25% of total TSS usage among different libraries deriving from a TSS at +4 (**Figure 12E**, brown TSS arrow and sequences) allowed study of positions -11 to -9 relative to this TSS (**Figure 12H**, top). As noted above, positions -8, -1, and +1 are major determinants for TSS efficiency. Interestingly, position -9 showed a relatively strong effect in our defined promoter context, which was not described in genome-wide analyses. At positions -4 to -2, we observed modulation of initiation efficiency, where in general C and/or G were preferred, and T was less preferred. This overall preference is consistent with an observation where individually mutating Ts at positions -4 to -2 relative to the -38 TSS of *ADH1* promoter to a C significantly increased usage of that TSS (Faitar et al., 2001). Though preferences at positions -7 to -5 are statistically significant (**Figure 12D**), contributions are much lower than other examined positions. Taken together, these results indicate that positions -9 to -8 and -4 to +1 are two major clusters contributing to TSS efficiency.

Experiments across species and the description of the canonical initiator element (Inr) suggest some sequence contributions from downstream positions relative to the TSS (Arkhipova, 1995; Basu et al., 2014; Deshpande and Patel, 2012; Faitar et al., 2001; Gleghorn et al., 2011; Javahery et al., 1994; Yarden et al., 2009; Zheng et al., 2011). To determine the impact of sequences downstream of the TSS, we examined motif enrichment of the most efficient -8 TSS variants, whose positions +1 to +9 are located in the randomized region (**Figure 12E**, green TSS arrow and sequences). We found an $A(/G)_{+2}G(/C)_{+3}G(/C)_{+4}$ enrichment for the top 10% most efficient TSS, but not for the next 10% most efficient TSS (**Figure 13**). These preferences are

consistent with early mutagenesis work in yeast, where A(+2) to C/T, G(+3) to T, or C(+4) to T substitutions decreased the utilization of a particular TSS, but A(+2) to G or T(+5) to C substitutions had minor effects (Faitar et al., 2001). This G(/C) enrichment might be related to higher stability of RNA-DNA hybridization or potential direct interaction with the Pol II active site (see Discussion).

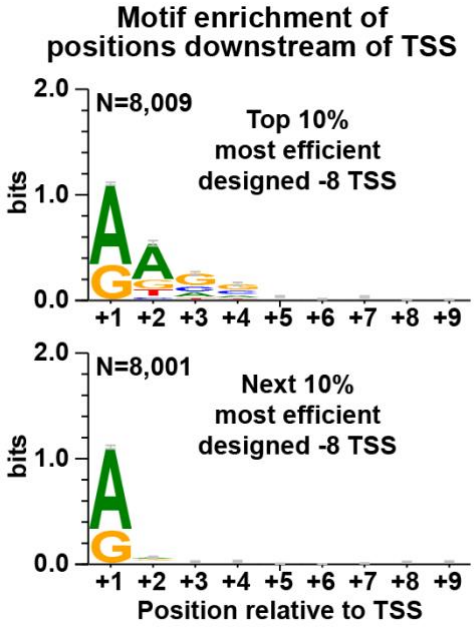


Figure 13 Sequence preference at TSS downstream positions

$A_{+2}G_{+3}G_{+4}$ motif enrichment is apparent for the top 10% most efficient designed -8 TSS. $A(/G)_{+2}G(/C)_{+3}G(/C)_{+4}$ motif enrichment was observed for the top 10% most efficient -8 TSSs but not for the next 10% most efficient TSSs. $A(/G)_{+1}$ enrichment observed for the top 20% most efficient TSSs is consistent with the +1R preference of TSS. Numbers (N) of variants assessed are shown.

Pairwise nucleotide-position dependencies have been observed in 5′ splice sites (Carmel et al., 2004; Roca et al., 2008; Wong et al., 2018). To investigate potential higher order sequence interactions *i.e.*, potential coupling between positions contributing to TSS efficiency, we examined

all possible pairwise interactions among positions -11 to +1 (**Figure 14, Figure 15**). "Coupling" would entail a base at one position determining the contribution or effect of a base at another position. We found evidence for coupling at multiple positions with the strongest coupling between positions -9 and -8. Here, an A at one position suppresses the preference of A at the other position, suggesting positional epistasis at each position where an A at one position might diminish the impact of an A at the other (**Figure 15A**). Using this -9/-8 interaction as an example, **Figure 14A** shows how coupling is detected and visualized. We calculated "centered relative efficiency" values for each base and position and visualized as a heatmap (**Figure 14).** We found that evident interactions were mainly observed at neighboring positions (**Figure 14B**), especially within positions -9 to -7 and within positions -5 to -2. In addition to its interaction with position -9 described above, position -8 was also observed to interact with position -7 (**Figure 15B**). Different from the suppression effect of As at positions -9 and -8, the presence of A at position -8 increased the relative preference of A or G compared to C or T at position -7. Related to the -8/-7 interaction, a direct contact between template strand bases at positions -7/-8 and GTF TFIIB B-reader residues was observed in a Pol II-TFIIB complex structure (Sainsbury et al., 2013). Together, these observations suggest that positions with observed interactions might work together in TSS selection (see Discussion).

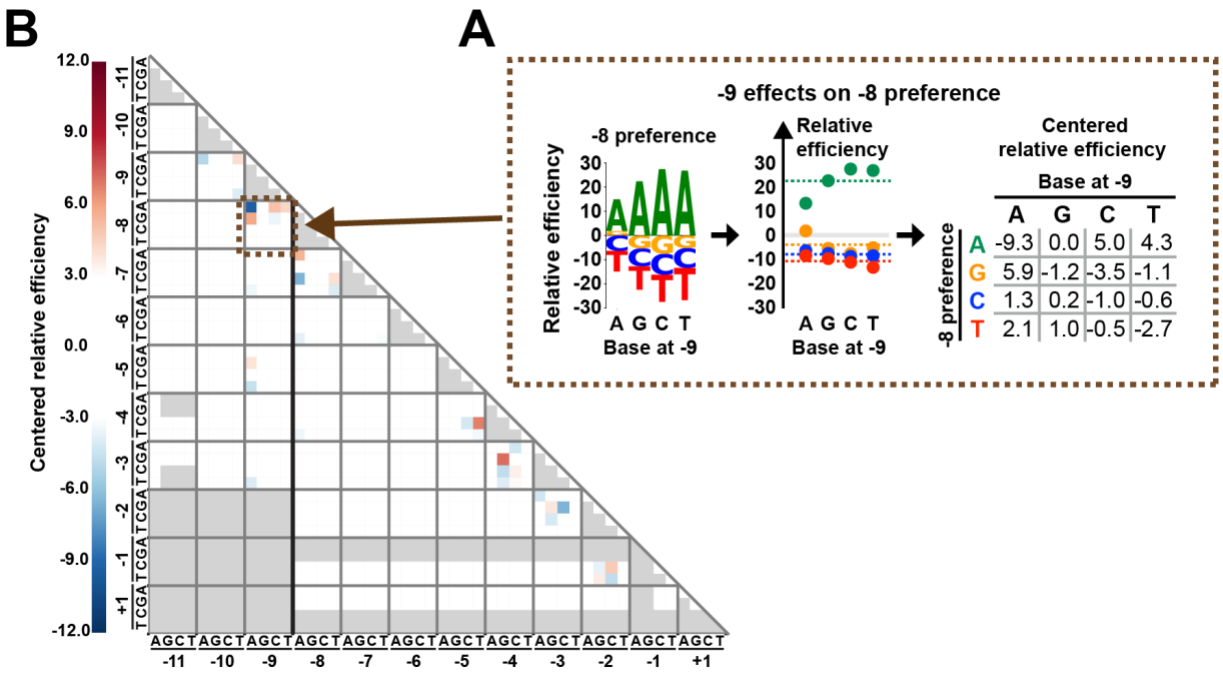


Figure 14 Sequence interaction between positions during TSS initiation

(A) Schematic illustrating how sequence interaction between positions is calculated and visualized as a heatmap in **B**. Using -9/-8 positions as an example, relative efficiencies at position -8 when different bases were present at position -9 were calculated. Next, relative efficiencies of each base were centered based on the mean of all relative efficiencies of a particular base. Finally, the centered relative efficiencies matrix was visualized as a heatmap to represent the interaction between examined positions. **(B)** Sequence interactions are mainly observed at neighboring positions. Red and blue indicate positive and negative interactions, respectively. Missing values are shown in grey. Interactions related to positions -11 to -9 were calculated using datasets of designed +4 TSSs deriving from "AYR", "BYR", and "ARY" libraries. Other interactions were calculated using datasets of designed +1 TSSs deriving from "AYR" and "BYR" libraries.

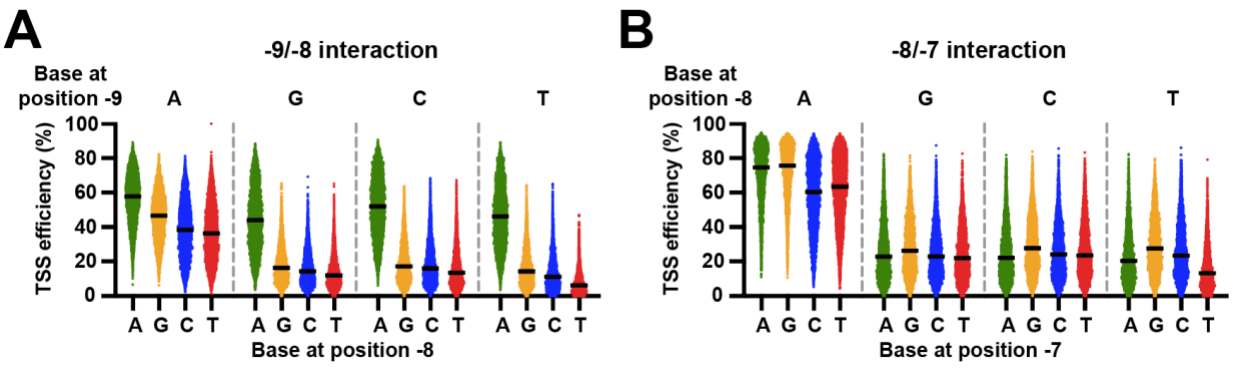


Figure 15 Sequence interactions of -9/-8 and -8/-7

(A) An A at position -9 results in different sequence preferences at position -8. The dataset of designed +4 TSSs deriving from "AYR", "BYR", and "ARY" libraries was used to detect -9/-8 interaction. All variants were divided into 16 subgroups defined by bases at positions -9 and -8 relative to designed +4 TSS, and then their TSS efficiencies were plotted. Lines represent median values of subgroups. **(B)** An A at position -8 results in different sequence preferences at position -7. The dataset of designed +1 TSSs deriving from "AYR" and "BYR" libraries was used to detect -8/-7 interaction. Calculations were the same as the -9/-8 interaction described in **A**.

2.2.3 Pol II mutants alter TSS efficiency for all possible TSS motifs while showing selective effects at +1

Pol II mutants were observed previously to change apparent specificity for A_{-8} versus B_{-8} (non-A) TSSs in opposite directions depending on the Pol II defect as determined by genomic TSS usage analysis (Qiu et al., 2020). As we noted in our prior work, this result could be a consequence of Pol II or GTF mutants shifting TSSs upstream or downstream coupled with uneven distributions of TSS motifs within genomic promoters, and the ability of a A_{-8} to function as an upstream R_{+1} TSS that might lack its own A_{-8}. To determine how Pol II catalytic activity affects TSS selection in our controlled promoter context, we measured effects on TSS efficiency in our promoter variant libraries between WT and Pol II mutants (**Figure 16, Figure 17**). We first investigated the overall effects of Pol II mutants on TSS usage (**Figure 17A**). As observed across the genome, Pol II GOF

mutants (G1097D and E1103G) shifted TSS usage upstream in aggregate for all libraries compared to WT, whereas Pol II LOF mutants (F1086S and H1085Q) shifted TSS usage downstream in aggregate across all libraries. We observed high reproducibility across biological replicates (**Figure 17B, C**). Therefore, we aggregated reads from biological replicates and employed cutoffs for total reads and promoters with variance above a threshold determined by the coefficient of variation (**Figure 17D**). Clustering analysis for efficiencies of major TSSs among all strains showed that Pol II WT and mutant classes (GOF and LOF) could be separated into groups based on mutant class (**Figure 17B**), consistent with Pol II mutant classes being distinguishable by a variety of profiling methods (Braberg et al., 2013; Qiu et al., 2016; Qiu et al., 2020). In total, our analysis allowed examination of >900,000 TSS sequences for each Pol II mutant (**Figure 16A**).

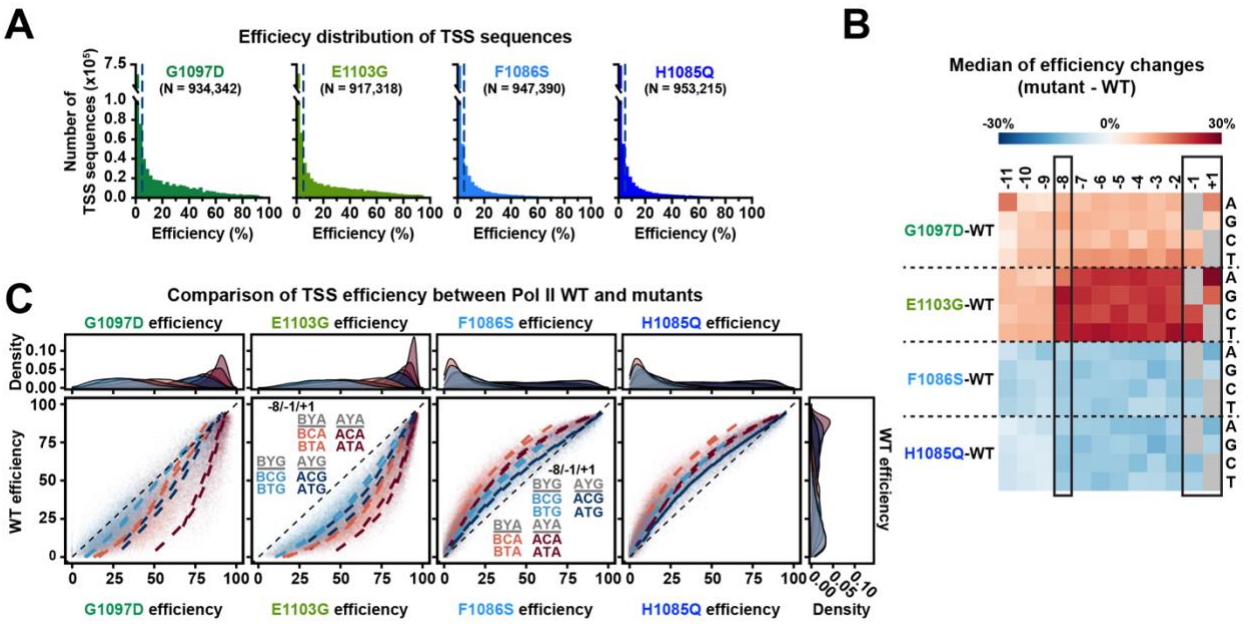
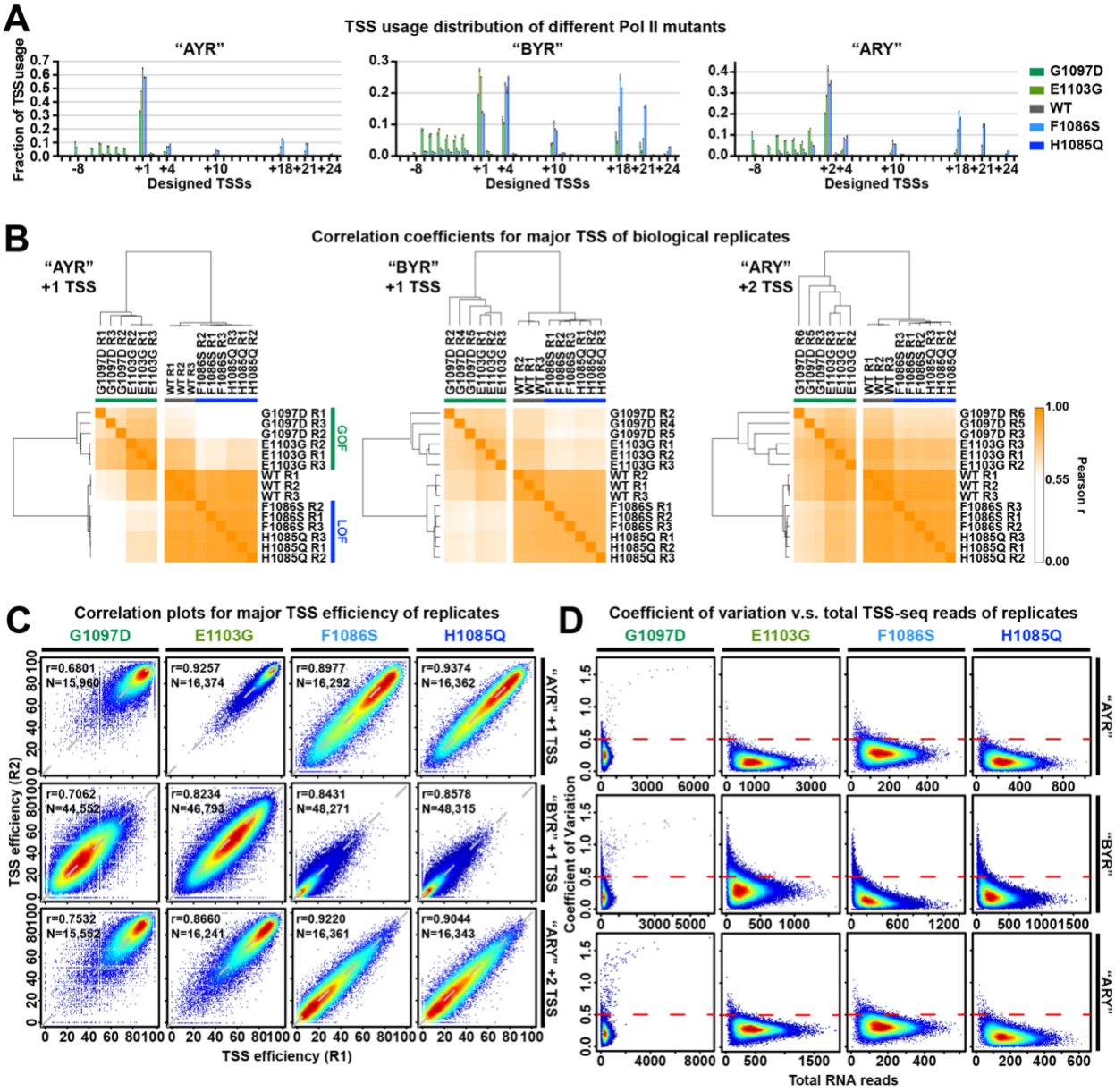
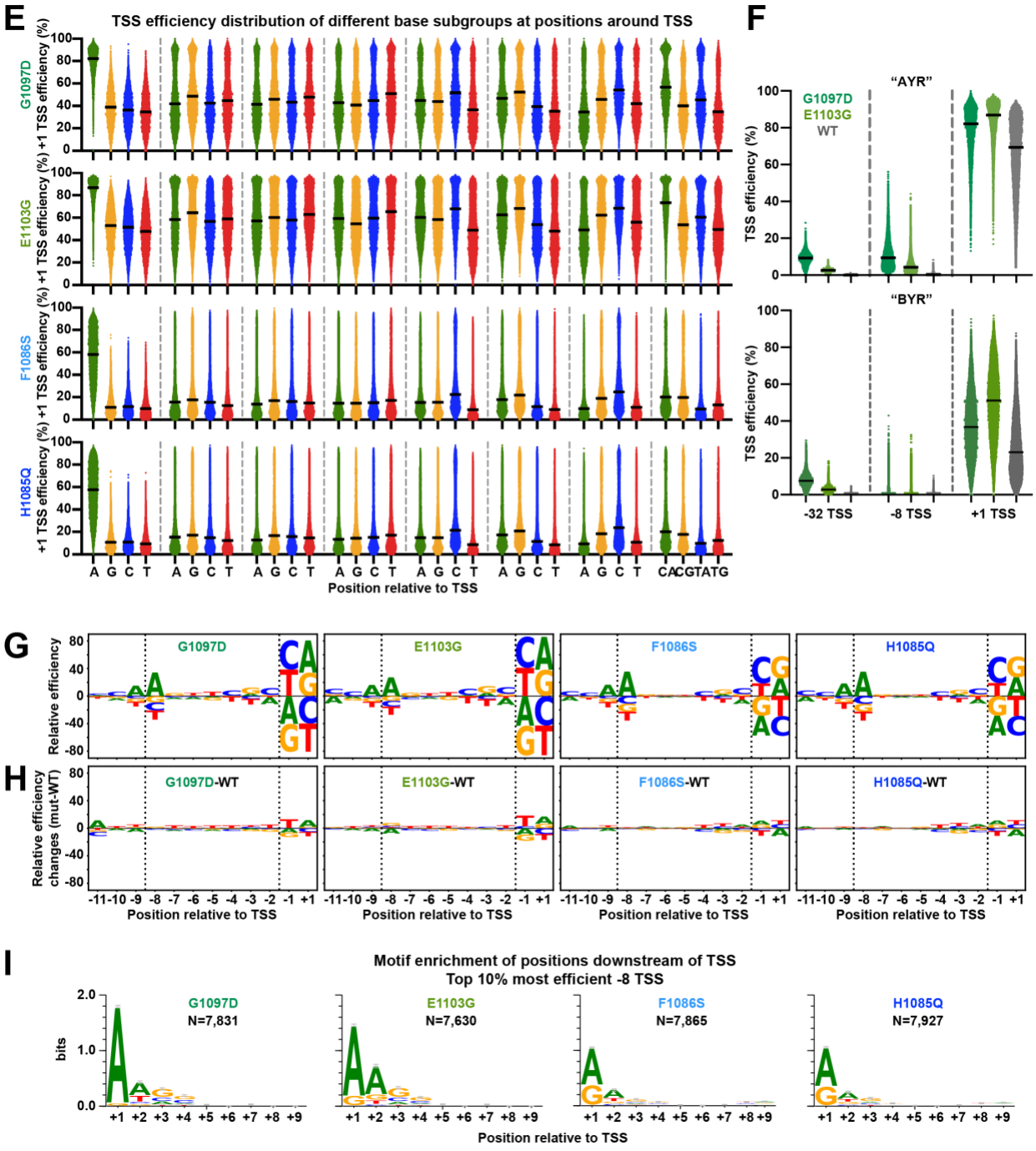


Figure 16 Pol II mutants alter TSS efficiency for all possible TSS motifs while showing selective effects for base at +1.

(A) Histograms showing the distribution of measured efficiencies for all designed -8 to +4 TSSs of all promoter variants deriving from "AYR", "BYR", and "ARY" libraries in Pol II mutants. Dashed lines mark the 5% efficiency cutoff used to designate a TSS as active. Total numbers of TSS sequences are shown. **(B)** Pol II mutants alter TSS

efficiencies across all motifs, corresponding to the direction of change to Pol II catalytic activity *in vitro*. TSS efficiency changes for each TSS variant were first determined by subtracting WT efficiency from Pol II mutant efficiency. The medians of efficiency changes for variant groups with indicated bases at each position relative to TSS were then calculated and illustrated in a heatmap. Positive (red) values indicate Pol II mutants increased overall efficiency while negative (blue) values indicate decreased overall efficiency. **(C)** WT TSS efficiency for TSS variants divided into motif groups are plotted for mutant TSS efficiencies for the same TSS groups. The eight possible groups of TSSs for $A/B_{-8}C/T_{-1}A/G_{+1}$ motifs were plotted and curve fit. Histograms show the density of variants within each -8/-1/+1 subgroups. As to position -8, A_{-8} containing motifs show higher efficiency than B_{-8} containing motifs in both Pol II GOF (G1097D and E1103G) and LOF (F1086S and H1085Q) mutants (A_{-8} motifs: maroon and blue vs B_{-8} motifs: light coral and light blue). This is consistent with the proposed function of -8A to retain TSSs longer in the Pol II active site during scanning. This means that -8A may boost the positive effects of GOF mutants, therefore Pol II GOF mutants showed greater effects on A_{-8} motifs compared to B_{-8} motifs. In contrast, -8A compensates for active site defects of LOF mutants, therefore Pol II LOF mutants showed reduced effects on A_{-8} motifs compared to B_{-8} motifs. Both GOF and LOF mutants show reduced effects on G_{+1} motifs relative to A_{+1} motifs (G_{+1} motifs: light coral and maroon vs A_{+1} motifs: light blue and blue).


A
TSS usage distribution of different Pol II mutants
"AYR"
"BYR"
"ARY"
Fraction of TSS usage
Designed TSSs
G1097D
E1103G
WT
F1086S
H1085Q
B
Correlation coefficients for major TSS of biological replicates
"AYR" +1 TSS
"BYR" +1 TSS
"ARY" +2 TSS
GOF
LOF
Pearson r
C
Correlation plots for major TSS efficiency of replicates
TSS efficiency (R2)
TSS efficiency (R1)
D
Coefficient of variation v.s. total TSS-seq reads of replicates
Coefficient of Variation
Total RNA reads


E
TSS efficiency distribution of different base subgroups at positions around TSS
G1097D
E1103G
F1086S
H1085Q
+1 TSS efficiency (%)
A G C T A G C T A G C T A G C T A G C T A G C T CACGTATG
Position relative to TSS
F
"AYR"
"BYR"
G1097D
E1103G
WT
TSS efficiency (%)
-32 TSS
-8 TSS
+1 TSS
G
H
Relative efficiency
Relative efficiency changes (mut-WT)
G1097D-WT
E1103G-WT
F1086S-WT
H1085Q-WT
-11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1
Position relative to TSS
I
Motif enrichment of positions downstream of TSS
Top 10% most efficient -8 TSS
bits
N=7,831
N=7,630
N=7,865
N=7,927
+1 +2 +3 +4 +5 +6 +7 +8 +9
Position relative to TSS

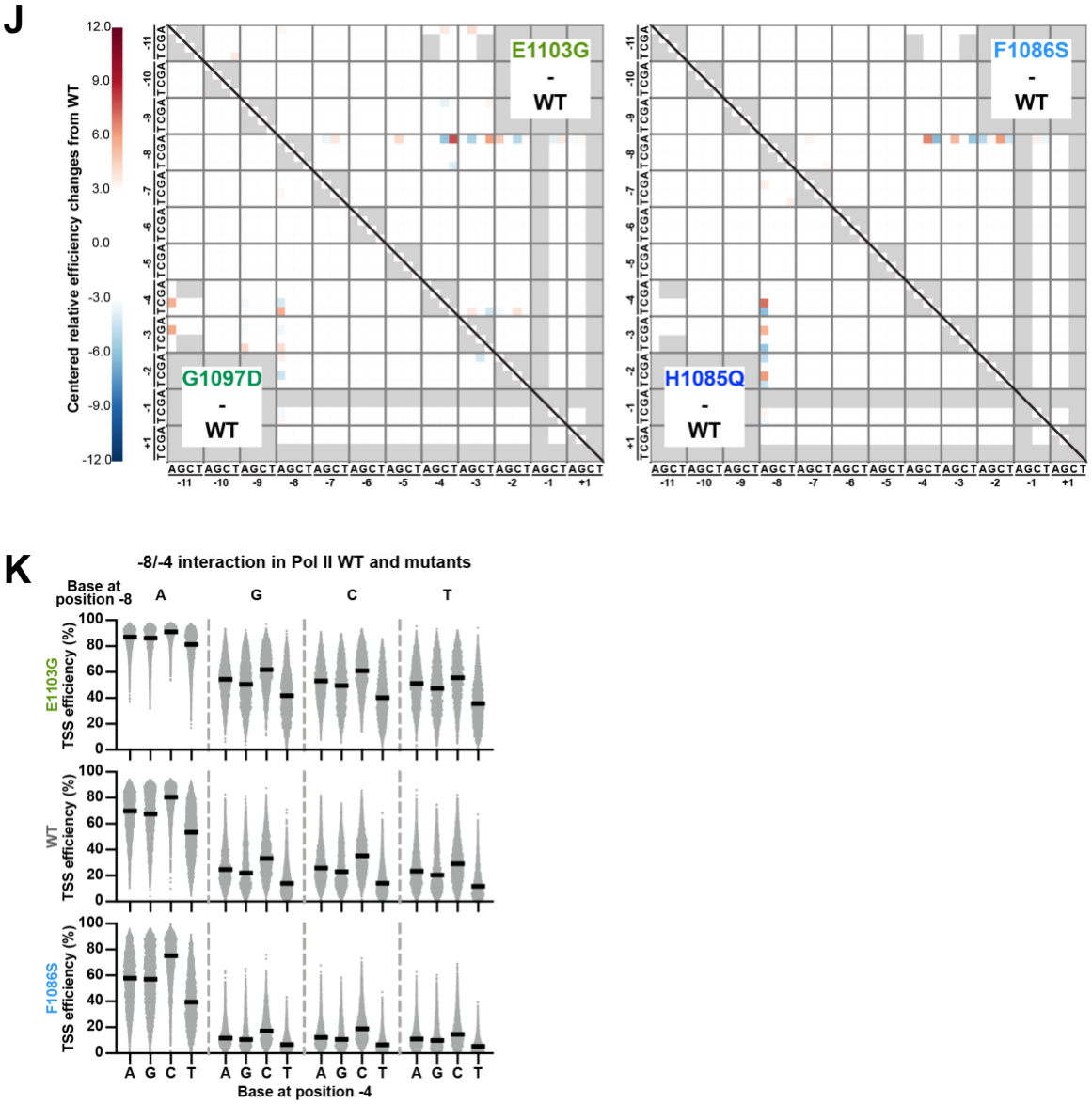


Figure 17 Pol II activity mutant effects on TSS selection

(A) TSS usage distributions at designed -10 to +25 TSSs in different promoter variant "AYR", "BYR", and "ARY" libraries. Pol II GOF and LOF mutants shifted TSS usage upstream or downstream relative to WT, respectively. Dots represent three biological replicates. Bars are mean +/- standard deviation of the mean. **(B)** Hierarchical clustering of Pearson correlation coefficients of TSS efficiencies for major TSSs (designed +1 TSS for "AYR" and "BYR" libraries, +2 TSS for "ARY" library) for three biological replicates for WT or mutant Pol II illustrated as a heatmap. **(C)** Example correlation plots of TSS efficiency of major TSSs of promoter variants between representative biological replicates. Pearson r and the number of compared variants are shown. **(D)** Plots of the Coefficient of Variation (CV) versus the total RNA reads for three yeast replicates in Pol II mutants. The red dashed lines mark the CV = 0.5 cutoff,

which was chosen as an arbitrary cutoff for variants showing reasonable reproducibility across three biological replicates. G1097D replicates contain outliers because these slow growing strains are susceptible to genetic suppressors. These outliers are filtered out using our CV cutoff because of high CVs caused by suppressor only existing in one of three biological replicates. **(E)** TSS efficiency distributions of designed +1 TSSs of Pol II mutants for base subgroups at individual positions relative to +1. Identical analysis as in **Figure 12D** for WT was performed for Pol II mutant libraries. **(F)** Pol II GOF G1097D showed a greater increase in efficiency than GOF allele E1103G at upstream TSSs (designed -32 and -8 TSSs), while E1103G showed stronger effects at designed +1 TSS than G1097D. This may indicate that where Pol II is in the scanning process may affect efficiency. **(G)** Pol II initiation sequence preference in Pol II mutants. Identical analysis as in **Figure 12H** for WT was performed for Pol II mutant libraries. Sequence logos were generated using multiple datasets and the dashed lines indicate divisions in datasets used to generate them. Specifically, preferences at positions -11 to -9 were generated using datasets of designed +4 TSSs deriving from "AYR", "BYR", and "ARY" libraries. Preferences at positions -8 to -2 were generated using datasets of designed +1 TSSs deriving from "AYR" and "BYR" libraries. Preferences at positions -1 and +1 were generated using datasets of designed +1 TSSs deriving from "AYR" and "ARY" libraries. **(H)** Pol II catalytic mutants do not change overall sequence specificity. Relative efficiency changes of mutants were calculated by subtracting relative efficiency in WT from that in mutants, followed by being visualized as sequence logos. The apparent changed preference at positions -1 and +1 could be biased by a ceiling effect for $Y_{-1}R_{+1}$ and a floor effect for $R_{-1}Y_{+1}$. This means that TSS efficiencies of $Y_{-1}R_{+1}$ containing variants have much less room to be increased by GOF mutants compare to $R_{-1}Y_{+1}$ containing variants. In contrast, TSS efficiencies of $Y_{-1}R_{+1}$ containing variants have much more room to be decreased by LOF mutants compared to $R_{-1}Y_{+1}$ containing variants. **(I)** Motif enrichment for the top 10% most efficient -8 TSSs for Pol II mutants. Identical motif enrichment analysis as in **Figure 13** top panel for WT was performed to Pol II mutant libraries. Numbers (N) of variants assessed are indicated. **(J)** Pol II mutants do not strongly affect interactions between promoter positions. Heatmaps show Pol II mutant effects on sequence interactions between promoter positions, as determined by subtracting centered relative efficiency in WT from that in mutants. The apparent altered -8A related interactions are biased by different levels of innate TSS efficiency and ceiling or floor effects, which are addressed in **K**. **(K)** Pol II mutants do not likely alter the -8/-4 interaction. Identical interaction analysis as in **Figure 15A** for the -9/-8 interaction in WT was performed to positions -8 and -4 using the dataset of designed +1 TSSs deriving from "AYR" and "BYR" libraries in WT, Pol II E1103G or F1086S libraries. No obvious interactions

were observed between positions -8 and -4 in WT (**Figure 14B**). The consistent relative preferences at position -4 (C > A ≥ G > T) in WT and Pol II mutants suggested Pol II mutants did not cause novel -8/-4 interaction. Instead, the apparent interaction (changed centered relative efficiency observed in **J**), was due to range effects, namely compression of efficiency at the low and high ends of the scale.

We first calculated TSS efficiency changes from WT for each TSS variant and then used the median values of base subgroups at each position to indicate effects (**Figure 16B**). Although slight apparent selective effects for specific sequences could be observed (addressed below), Pol II mutants showed directional effects on TSS efficiencies across all TSS motifs, depending on predicted changes to Pol II catalytic activity (**Figure 16B, Figure 17E**). Specifically, Pol II hyperactive mutants (G1097D and E1103G) increased overall efficiency at all TSS sequences, whereas Pol II hypoactive mutants (F1086S and H1085Q) decreased overall efficiency for all TSS sequences. The same direction of effects on preferences for A_{-8} and B_{-8} containing TSSs argues against the hypothesis that observed divergent effects on the usage of A_{-8} and B_{-8} containing motifs from genomic promoters were derived from changes to innate preference for TSS motifs. Therefore, the apparent changed selectivity for A_{-8} versus B_{-8} containing TSSs observed in genomic analysis is likely an indirect effect of Pol II mutants, resulting from polar shifts in TSS distribution relative to the uneven underlying TSS motif distributions. We did observe selective effects of Pol II mutants on TSS efficiencies for specific sequences beyond the uniform polarities of changes observed across all sequences. Specifically, both Pol II GOF and LOF mutants showed apparent reduced effects on A_{-8} compared to non-A_{-8} containing motifs. Additionally, effects on G_{+1} containing motifs compared to A_{+1} motifs appeared reduced as well (**Figure 16B**). Some of these apparent differences could be more apparent than real due to different levels of innate TSS efficiency and range effects. Range effects, such as “ceiling” or “floor” effects, arise from

compression in signal range; for example, TSS efficiency cannot be better than 100% or worse than 0%. To visualize this, efficiencies of $N_{-8}Y_{-1}R_{+1}$ variants in mutants were plotted based on their WT efficiency and colored based on -8/-1/+1 subgroups to control for confounding variable of different average efficiencies for particular motifs (**Figure 16C**). We observed two independent and additive patterns related to A_{-8}/B_{-8} and A_{+1}/G_{+1}, discussed below.

First, A_{-8} containing TSSs, including $A_{-8}Y_{-1}A_{+1}$ and $A_{-8}Y_{-1}G_{+1}$, showed higher efficiency than B_{-8} TSSs in aggregate for both Pol II GOF and LOF mutants across all TSS efficiencies. It also appears the relatively smaller Pol II GOF effects on A_{-8} containing TSSs observed in **Figure 16B** are likely due to a "ceiling" on TSS efficiency, where many A_{-8} containing TSSs that are already highly efficient have little room for further increase in efficiency (**Figure 12B, D**). Second, A_{+1} containing TSSs, including $A_{-8}Y_{-1}A_{+1}$ and $B_{-8}Y_{-1}A_{+1}$, showed higher efficiencies than G_{+1} containing TSSs in GOF but lower efficiencies in LOF mutants (**Figure 16B, C**). This provocative result supports potential Pol II active site control over initiation NTP preference where Pol II active site sensing of NTP levels can control initiation decisions in meaningful ways (Jenks et al., 2008; Kuehner and Brow, 2008). Alternatively, differential usage of G_{+1} versus A_{+1} sites could reflect possible alterations to GTP/ATP ratios in Pol II mutants (see Discussion).

We note that Pol II GOF mutant G1097D, our strongest hyperactive mutant *in vitro* with a severe growth defect *in vivo*, generally increased efficiency to a lesser extent than E1103G (**Figure 16B, Figure 17E, F**). That was unexpected because G1097D showed a higher *in vitro* elongation rate than E1103G and stronger effects on TSS shifts at *ADH1*, *GAL1*, and *IMD2* (Kaplan et al., 2012) and genome-wide (Qiu et al., 2020), along with its aforementioned stronger growth defects. Consistent with its usually stronger phenotypes, we observed greater far upstream TSS usage from our promoters in G1097D than E1103G (**Figure 17A**). To reconcile these observations, we

speculate that the position of TSS within a promoter or scanning window might also affect TSS efficiency. This means that while we observe upstream TSS shifting due to increased overall efficiency, there may be complexities in the scanning process such that downstream TSSs are affected differently or appear to have decreased efficiency as initiation approaches background levels. Evidence supporting this hypothesis is, at upstream TSSs, such as designed -32 TSS and -8 TSS, G1097D showed higher efficiency than E1103G (**Figure 17F**). At genomic promoters, GOF mutants were observed to increase TSS efficiency of TSSs upstream of promoter "median" TSSs but decrease efficiency of downstream TSSs relative to WT (Qiu et al., 2020).

We did not observe strong effects on sequence preference at positions upstream of TSS or sequence interaction in Pol II mutants (**Figure 17G, H, J, K**), suggesting these attributes do not have specific interactions with altered Pol II activity. However, we did observe different base enrichment at position +2 of the most efficient -8 TSSs in Pol II mutants (**Figure 17I**) compared to that in WT (**Figure 13**), suggesting the Pol II active site might interact with this position. Specifically, we observed differences in A/G enrichment at position +2 between WT and Pol II mutants, especially GOF mutants. Compared to WT (**Figure 13**), G1097D decreased +2G enrichment, whereas E1103G increased +2A enrichment. These Pol II mutant alterations on sequence enrichment indicate that Pol II catalytic preferences might be impacted by bases at position +2 and that G1097D and E1103G might alter Pol II activity in allele-specific ways.

2.2.4 Learned initiation preferences are predictive of TSS efficiencies at genomic promoters

To ask how sequence determinants identified here relate to natural promoters, we compared our library-defined sequence efficiencies to TSS efficiencies observed at genomic promoters (**Figure 18**). To limit potentially confounding factors for genomic promoters, we focused on a

single “median” TSS for each promoter in a defined set of promoter windows. The median TSS is defined as the promoter position containing the 50th percentile of reads within each promoter window (Qiu et al., 2020). We found that Pol II sequence preference at positions around median TSSs was mostly consistent with what we observed in our libraries (**Figure 12H, Figure 18A**). Efficient genomic TSSs appear enriched for A at positions -7 to -5. The A-richness at positions between -10 to -3 and +5 to +10 has been noted in previous studies from our lab (Qiu et al., 2020) (**Figure 18B**) and others (Lu and Lin, 2019; Malabat et al., 2015; Zhang and Dietrich, 2005). However, As at positions -7 to -5 appeared neutral in our promoter libraries (**Figure 12H**). The observed A-richness in the genome could reflect selection *in vivo* for additional promoter properties, such as providing an easily meltable DNA region, lower nucleosome occupancy or reflect a context dependent role not reflected by our promoter libraries.

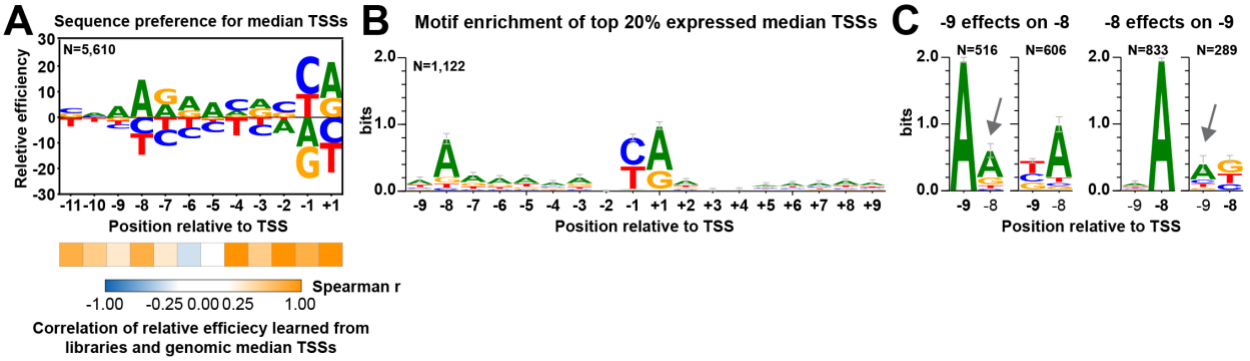


Figure 18 Learned initiation preferences are predictive of TSS efficiencies at genomic promoters

(A) Sequence preference determined from genomic median TSSs are congruent with library determined TSS preferences, except there is genomic enrichment for A at positions -7 to -5. Sequence context and TSS efficiency of genomic median TSSs were extracted from genomic TSS-seq data (GSE182792) (Zhao et al., 2021). Calculation and visualization were as performed for promoter variant libraries. The number (N) of genomic median TSSs examined is shown. Statistical analyses by Spearman’s rank correlation test between relative efficiency at individual positions learned from promoter variant libraries and genomic median TSSs are shown beneath the sequence logo. **(B)** A-richness upstream and downstream of TSS is observed for highly expressed median TSSs. The number (N) of analyzed

median TSSs is shown. **(C)** Having an A at either position -9 or -8 reduces the enrichment of A at the other position. The top 20% expressed median TSSs were divided into subgroups based on bases at position -9 or -8. Motif enrichment analysis was individually performed on subgroups. Numbers (N) of median TSSs within each subgroup are shown.

We find that the interaction between -9A and -8A discovered in our libraries is also reflected in genomic promoters (**Figure 18C**). We first grouped the median TSSs from the top 20% expressed promoters based on the base at position -9 or -8 and then examined sequence enrichment at the other position. We observed that -9A decreased the enrichment of -8A relative to that when position -9 is not A (**Figure 18C**, left). Moreover, when A was absent from the position -8, much higher enrichment for -9A was observed (**Figure 18C**, right). Together, consistent with -9/-8 interaction discovered in our controlled context, in the genome the presence of an A at either position -9 or -8 suppresses the enrichment of A at the other position within the group of most efficiently used median TSSs. These results suggest that -9A may function in a similar fashion as -8A, but that -8A is more efficient and therefore has been evolutionarily favored (see Discussion).

2.2.5 Regression modeling identifies key DNA sequences and interaction for TSS selection regulation

We have found that DNA sequences around the TSS not only additively but also interactively contribute to TSS efficiency. To quantitatively identify key features (sequences and interactions) for TSS efficiency, we compiled datasets derived from all libraries and predicted TSS efficiency from sequence information by logistic regression coupled with a forward stepwise selection strategy (**Figure 19A-F, Figure 20A, B**). We first compiled datasets generated from

designed -8 to +2 and +4 TSSs deriving from "AYR", "BYR", and "ARY" libraries (**Figure 19A**) and split data into training (80%) and test (20%) sets. For each variant, its sequences at positions -11 to +9 were extracted as potential predictors for TSS efficiency. We then used a forward stepwise strategy with a 5-fold Cross-Validation (CV) to select robust features (predictors). By evaluating model performance with R^2, sequences at nine positions (positions -9 to -7 and -4 to +2) and one interaction (-9/-8 interaction) were identified as robust features and selected for final modeling (**Figure 19B**). It is worth noting that models with as few as three features – sequences at positions -9, -8, and -1 (or +1) – could explain 74.10% of TSS efficiency variation.

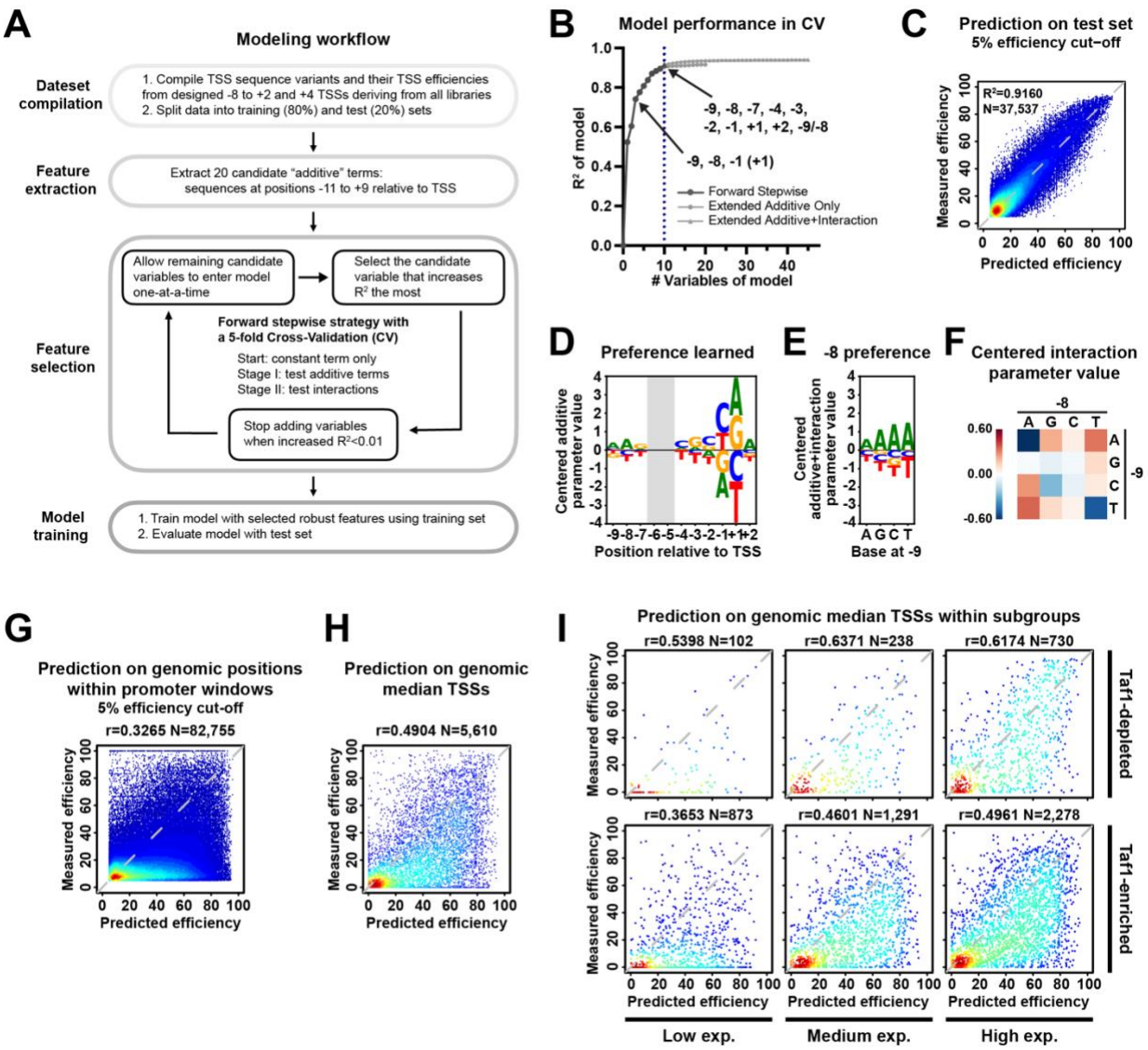


Figure 19 Logistic regression models of TSS sequence function

(A) Overview of modeling process. (1) Variants including designed -8 to +2 and +4 TSSs deriving from "AYR", "BYR", and "ARY" libraries with available TSS efficiency were pooled for modeling. (2) Sequences at positions -11 to +9 relative to TSS of each variant were extracted. (3) To identify robust features, a forward stepwise selection strategy coupled with a 5-fold Cross-Validation (CV) for logistic regression was used. Data were randomly split into training (80%) and test (20%) sets. The training set was used for a stepwise regression approach that starts from a model with a constant term only and adds variables that improve the model the most one at a time until a stopping criterion is met. In stage I, additive terms (sequences at positions -11 to +9) were tested. In stage II, interactions between positions selected in stage I were tested. Model performance was evaluated with R^2. The stopping criterion for adding additional variables was an increase $R^2 < 0.01$. (4) A logistic regression model containing selected robust

features was trained with the training set and then evaluated with the test set. **(B)** Regression modeling identifies key DNA sequences and interactions contributing to TSS efficiency. Nine additive parameters plus one interaction were selected for the final model. Dots represent average R^2 obtained in a 5-fold CV strategy for logistic regression models using different numbers of features. The black line with SD error bars represents models with the best performance under a certain number of predictors. Note that a model with as few as three additive parameters could explain 74.10% of TSS efficiency variation in our WT libraries. **(C** to **F)** Good performance of model including sequences at nine positions and the -9/-8 interaction indicates that TSS efficiency in our libraries is mainly regulated by the included features. (**C**) A scatterplot of comparison of measured and predicted efficiencies of TSS sequence variants in the test set, with a 5% efficiency cut-off. Model performance R^2 on the entire test set and the number (N) of data points shown in the plot are shown. (**D**) A sequence logo of centered additive parameters. The coefficients for bases at a particular position were centered and visualized as a sequence logo. (**E**) A sequence logo showing learned preference at position -8 when different bases exist at position -9, with -9/-8 interaction included. The -9/-8 interaction parameters were added to corresponding additive coefficients for bases at position -8. The additive plus interaction parameters were then centered and visualized as sequence logos. (**F**) A heatmap of centered parameters for the -9/-8 interaction illustrating how bases at one position affect preference at another position. **(G** to **H)** Efficiency prediction for positions within known promoter windows in WT shows overall over-prediction. Scatterplots of comparison of measured and predicted TSS efficiencies of all positions (with a 5% efficiency cut-off) (**G**) or median TSSs (**H**) within 5979 known genomic promoter windows (Qiu et al., 2020) with available measured efficiency. **(I)** Model shows better performance on Taf1-depleted promoters and promoters with medium to high expression. Scatterplots of comparison of measured and predicted TSS efficiency of median TSSs subgrouped by promoter classes and expression levels. Expression levels of genomic promoters are defined based on their total TSS-seq reads in the promoter window in the examined datasets: low, [0, 200); medium, [200, 1000); high, [1000, max). Pearson r and the number (N) of compared variants are shown.

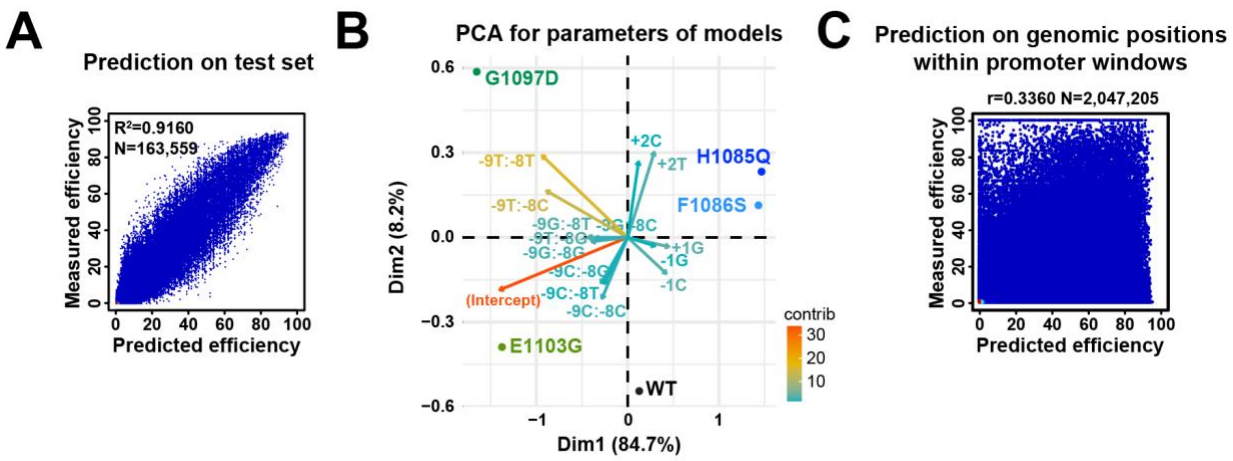


Figure 20 Efficiency prediction on the test set and genomic positions and PCA for parameters of models

(A) A scatterplot of comparison of measured and predicted efficiencies of all TSS sequence variants in the test set. Model performance R^2 on the entire test set and the number (N) of data points shown in the plot are shown. **(B)** PCA analysis for parameters of models trained by WT and Pol II mutant datasets. The top 15 contributing variables are shown. GOF and LOF mutants were separated from WT by the 1st principal component. GOF G1097D and E1103G were further distinguished by the 2nd principal component by additional position +2 information, which is consistent with results in **Figure 17I**, where G1097D and E1103G differentially altered +2 sequence enrichment. **(C)** A scatterplot of comparison of measured and predicted TSS efficiencies of all positions within 5979 known genomic promoter windows (Qiu et al., 2020) with available measured efficiency. Pearson r and the number (N) of compared variants are shown. Most promoter positions (82%, 1,678,406 out of 2,047,205) showed no observed efficiency, which is expected because TSSs need to be specified by a core promoter and scanning occurs over some distance downstream.

The final model containing the most predictive features explained 91.60% of the variance in TSS efficiency for the WT test set (20% of the total dataset) (**Figure 19C, Figure 20A**). We next asked whether the features learned by modeling using compiled datasets were consistent with our previous sequence preference analysis using selected representative datasets with the most randomized bases. We centered additive variable values and visualized them as a sequence logo (**Figure 19D**). First, as expected, positions -1 and +1 were the major predictors; however, the

influence of the -8A did not appear as strong as in our previous preference analysis. We suspected this might be because the -9/-8 interaction contribution was not included. After adding the -9/-8 interaction term, we observed an emergence of the position -8 as an influential predictor (**Figure 19E, F**), which also emphasizes the contribution of the -9/-8 interaction. Second, and importantly, modeling confirmed the +2A preference observed in previous motif enrichment analysis using only the most efficient -8 TSS variants (**Figure 13**). Additionally, the result that the forward stepwise modeling selected sequences at position +2 for the final model further emphasized the importance of position +2. The impact of position +2 is also evident when performing Principal Component Analysis (PCA) to variables of models trained with WT or Pol II mutant datasets (**Figure 20B**). The fact that sequences at position +2 are top contributing variables in the 2nd principal component that distinguished G1097D and E1103G is in agreement with differentially altered +2 sequence enrichment by Pol II mutants (**Figure 17I**), suggesting position +2 preference is altered by Pol II activity changes, and that this position might work directly with the Pol II active site.

2.2.6 Sequence defines TSS efficiency within a wider promoter context during initiation by scanning

To evaluate the extent to which DNA sequence around a TSS contributes to TSS efficiency at genomic promoters, we compared the difference between observed and model predicted efficiencies of all positions within known promoter windows or within specific subgroups of known promoters (**Figure 19G-I, Figure 20C**). As expected, we found most promoter positions showed low or no observed efficiency and were over-predicted by sequence alone (**Figure 19G, Figure 20C**), because TSSs need to be specified by a core promoter and scanning occurs over

some distance downstream. Therefore, an individual potential TSS has additional attributes that may affect its efficiency beyond local sequence, such as distance from the site of PIC assembly. We therefore extracted only median TSSs purportedly within promoter scanning windows to ask how our sequence-based predictor functions on genomic TSSs (**Figure 19H**). We also separated median TSSs by promoter classes based on Taf1 enrichment (Rhee and Pugh, 2012), a proxy for the two main types of promoters in yeast, or promoter expression levels (**Figure 19I**). We observed good prediction performance at a wide range of TSSs, indicating sequence determinants identified in our limited promoter context contribute to TSS efficiency in genomic promoter contexts. We observed increased performance for higher expressed promoters (Pearson r increased from 0.37-0.54 to 0.46-0.64) and Taf1-depleted promoters (Pearson r increased from 0.37-0.50 to 0.54-0.64) (**Figure 19I**). Conceivably, highly expressed promoters may have evolved TSSs at optimal distances from core promoters, and therefore may be similarly sensitive to sequence effects. Alternatively, the use in our libraries of *GAL1* UAS and *SNR37* core promoter elements, where both *GAL1* and *SNR37* are highly-expressed Taf1-depleted promoters, may share sequence sensitivities for TSSs from related promoters.

2.3 Discussion

Individual TSS sequences are critically important for initiation output, but this has not been systematically determined for eukaryotic promoters. Changes to initiation factor activity may also alter initiation preferences due to functional if not physical interaction with DNA sequence. Studies examining TSS selection have been based on existing gene promoters *in vivo* and are subject to unknown biases or confounding variables of different promoter contexts. Here we

developed and employed Pol II MASTER to systematically investigate ~1 million TSS sequences in wildtype or Pol II mutant cells. This system allowed us to specifically and comprehensively study TSS efficiencies in initiation by promoter scanning, without confounding effects from other architectural features, such as variability in core promoter-TSS distances, differences in promoter identities or chromatin configurations that may obfuscate analyses of genomic TSSs. We find sequence variation at different positions around TSS considerably tunes initiation efficiency in a predictable way and these contributions are important for initiation efficiency at genomic promoters.

Combining results from this study and others, we suggest how TSS sequence contribution for TSS selection works (**Figure 21**). We propose that two major groups contribute to TSS selection: bases around TSS (actual initiating site) and bases around position -8. First, in promoter scanning, the TSS and adjacent bases interact with the Pol II active site, the 1st NTP or each other to facilitate stable binding of the 1st NTP and potentially the 2nd NTP to stimulate RNA synthesis. This would be in contrast to the concept of the initiator or downstream elements functioning as part of the TFIID or PIC binding site as has been proposed for higher eukaryotes (see (Luse et al., 2020; Vo ngoc et al., 2020) and references therein). As the universal initiating element, $Y_{-1}R_{+1}$ has been established to facilitate stable binding of the NTPs by RNA polymerases via base stacking between R_{-1} from the template DNA and the 1st purine NTP (Basu et al., 2014; Gleghorn et al., 2011). Positions upstream of the TSS, such as positions -4 to -2, might contribute to stabilizing template DNA via base stacking or physical interaction with initiation factors (Basu et al., 2014; Zhang et al., 2012).

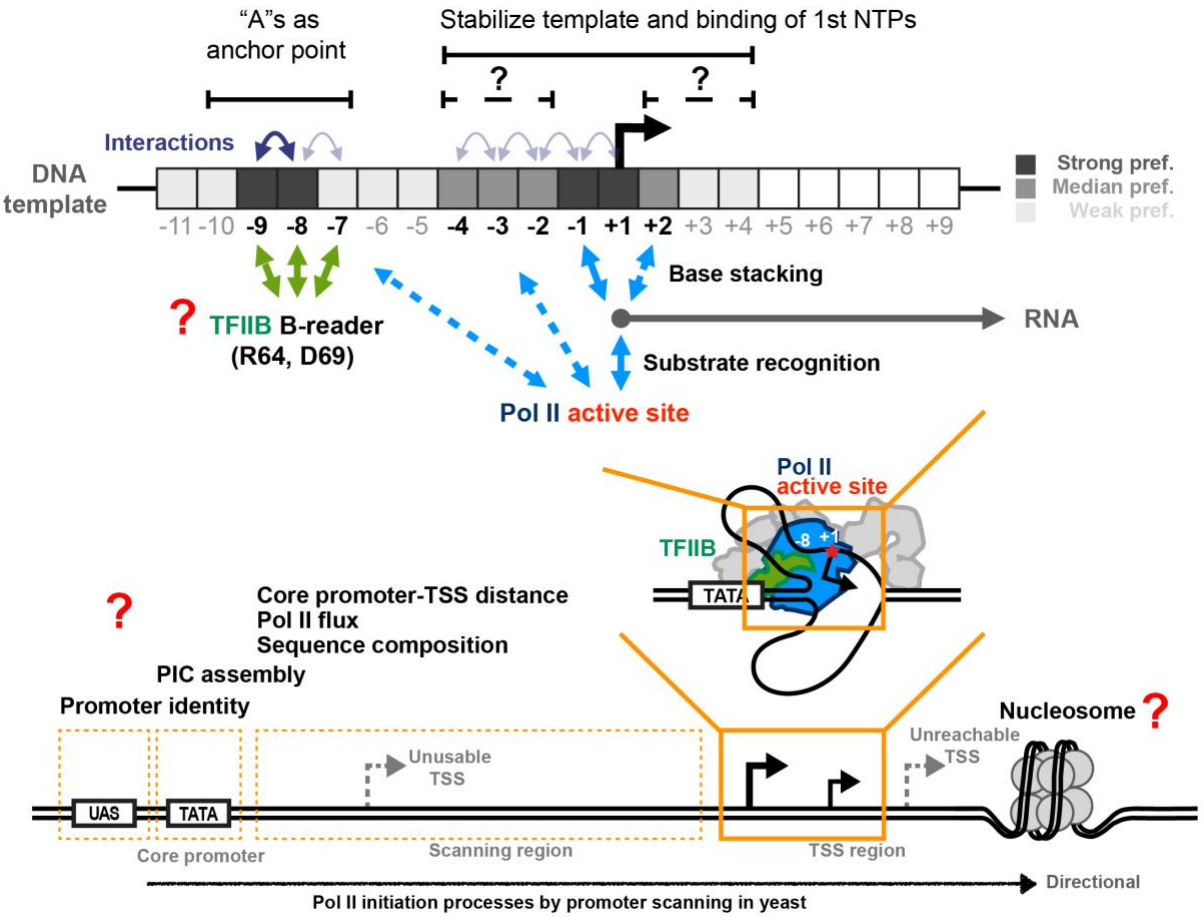


Figure 21 Model for TSS sequence preference regulated by multiple mechanisms.

The top panel shows determined contribution of sequence at positions around TSS and proposed corresponding mechanism by this study. Two major groups of positions around TSS contribute to TSS selection: bases around TSS (actual initiating site) and bases around position -8. The TSS and adjacent bases interact with the Pol II active site, the 1st NTP or each other to facilitate stable binding of the 1st NTP and thus stimulate RNA synthesis. -8 and -9 Ts on the template strand with an additional interaction between -8 and -7 template strand positions are proposed to serve as an anchor point interacting with the TFIIB B-reader domain allowing pausing of scanning and promotion of Pol II initiation at TSSs a fixed distance downstream. These preferences are reflected as As if the analysis is on the coding strand. Positions and interactions that were identified by regression modeling as robust features are labelled in bold. The bottom panel shows other architectural features involved in Pol II transcription initiation likely additionally contributing to TSS selection and initiation efficiency that will be accessible to Pol II MASTER analysis.

Our observation that Pol II mutants showed selective effects on base at position +1 (**Figure 16**) supports that position +1 functions through the Pol II active site and could suggest a mechanism for cellular state to regulate initiation via alteration of initiating base (ATP/GTP) ratios. We have observed that A_{+1} containing TSSs increase in efficiency more than G_{+1} containing TSSs in Pol II GOF mutants, while A_{+1} containing TSSs decrease in efficiency more than G_{+1} containing TSSs in Pol II LOF mutants. In other words, in both classes of Pol II mutants, G_{+1} containing TSSs appeared relatively buffered compared to A_{+1} containing TSSs. Two possible explanations could be envisioned. The first explanation is that differential preference for ATP versus GTP is directly affected by altered Pol II activity. The second is that differential effects on TSS efficiency for +1 A sites versus +1 G sites result from indirect effects on ATP/GTP ratio in cells. Such defects might result from altered synthesis of nucleotide synthesis-related genes, a number of which are themselves sensitive to Pol II activity (Braberg et al., 2013; Kaplan et al., 2012; Kuehner and Brow, 2008; Kwapisz et al., 2008; Malik et al., 2017; Thiebaut et al., 2008). More specifically, a hypothetical decreased GTP/ATP ratio in GOF mutants would suppress increased initiation efficiency of G_{+1} containing TSSs relative to A_{+1} containing TSSs, while a hypothetical increased GTP/ATP ratio in LOF mutants would compensate for a decrease in efficiency of G_{+1} containing TSSs. This explanation would be consistent by multiple observations of ATP- and GTP-related genes in Pol II mutants. In GOF mutants, the expression of GTP-related genes, *IMD2*, *IMD3*, and *IMD4*, is known to be defective (Braberg et al., 2013; Kaplan et al., 2012), while in Pol II LOF mutants, *IMD2* is constitutively active (Kaplan et al., 2012; Malik et al., 2017) and overall expression at the mRNA level for *IMD* genes is increased (Kaplan et al., 2012). These are consistent with differential effects on GTP relative to ATP synthesis as the products of the *IMD* genes (inosine monophosphate dehydrogenase) compete with the *ADE12*

product (adenylosuccinate synthase) for inosine monophosphate (IMP) in the synthesis of GMP and AMP precursors, respectively (Kanehisa and Goto, 2000; Rolfes, 2006; Strathern et al., 1982). Furthermore, Pol II GOF E1103G was determined to have reduced levels of guanine, which could have knock-on effects to GTP levels or reflect increased demand for guanine-related metabolites (Gout et al., 2017).

The positions +2 to +4 downstream of TSS might contribute to establishing NTP stability in the Pol II active site as well (**Figure 13**). In addition, we observed Pol II mutant effects on those preferences (**Figure 17I, Figure 20B**), suggesting these positions might function via directly interacting with the Pol II active site, as observed in transcription structures in other species (Basu et al., 2014; Gleghorn et al., 2011). First, T7-like single-subunit RNAP family showed a base-specific interaction between the 2nd NTP and a residue in the middle of the O helix, which was suggested to enhance the formation of the first phosphodiester bond (Gleghorn et al., 2011). Second, a structure of *de novo* transcription initiation complex in bacterial RNA Polymerase showed multiple interactions between the 2nd NTP and its β′ and β subunits, whose eukaryotic counterparts are Rpb1 and Rpb2 (Basu et al., 2014). Alternatively, the A/G-rich $A(/G)_{+1}A(/G)_{+2}G_{+3}G_{+4}$ motif might be related to translocation. A study from the Landick lab showed that when A/G comprised the RNA 3′ end, the RNAP active site favored the post-translocated state (Hein et al., 2011). If A/G similarly effects translocation state within the first four bases in Pol II initiation, synthesis of the first few bases might be promoted. Together, positions downstream of TSS might contribute to stabilizing or facilitating the first few NTPs adding.

Where TFIIB has been specifically implicated is in bases near position -8, where the TFIIB B-reader domain has been observed to directly interact in a structure of a yeast Pol II-TFIIB

complex (Sainsbury et al., 2013). Here, it is attractive to envision TFIIB functioning as an anchor point to pause the scanning process to promote Pol II initiation at a fixed distance downstream. Several observations support this proposed function. First, we detected sequence interaction between positions -8 and -7 (**Figure 14B, Figure 15B**). This is in line with the direct contact of -7T and -8T on the template strand and TFIIB B-reader R64 and D69 observed in a Pol II-TFIIB complex structure (Sainsbury et al., 2013), which has been proposed to hold TSSs in the Pol II active site longer during scanning. Second, we observed a strong interaction between positions -9 and -8, where the presence of an A at either position suppressed the preference of A at the other position (**Figure 14, Figure 15A**). This -9/-8 interaction was also evident when examining genomic median TSSs (**Figure 18C**). Taken together, we speculate that Ts around position -8 or -9 on the template strand and TFIIB may pause the scanning process to facilitate the usage of TSSs positioned 8 to 9 bases downstream. Moreover, we have shown that Pol II catalytic mutants alter TSS efficiencies across all TSS sequences, without showing alteration in preference for -8A (**Figure 16B, C, Figure 17E**), suggesting that Pol II catalytic activity is not responsible for -8 preference.

Whether or how DNA sequence surrounding the TSSs is involved in other promoter properties is another question. As at positions -7 to -5 were measured as neutral in our promoter libraries (**Figure 12H**), in contrast to the A-enrichment at highly expressed and focused genomic TSSs (**Figure 18A, B**) (Lu and Lin, 2019; Maicas and Friesen, 1990; Malabat et al., 2015; Qiu et al., 2020; Zhang and Dietrich, 2005). We speculate that observed A-richness around TSS functions through other evolved promoter properties. First, observed A-richness between positions -10 to -3, together with T-richness at further upstream core promoter region (Lubliner et al., 2013; Maicas and Friesen, 1990), provides an easily meltable region for DNA unwinding, perhaps facilitating

transcription initiation in specific contexts. Second, higher A/T content may function to lower nucleosome occupancy (Segal and Widom, 2009), because appropriately periodic G/C dinucleotides promote nucleosome occupancy (Lee et al., 2007; Peckham et al., 2007; Segal et al., 2006; Tillo and Hughes, 2009). However, the base composition switch in highly expressed promoters from T- to A-preponderance (Maicas and Friesen, 1990; Qiu et al., 2020; Wu and Li, 2010) indicates A-richness may have other roles depending on the characteristics of the sequence itself. Third, A-richness may be left over from the evolution of promoter scanning. A recent study of transcription initiation mechanism investigated 12 yeast species and proposed that during evolution an A-rich region upstream of TSS appeared first, then the specific -8A preference occurred (Lu and Lin, 2020). Therefore, the observed A preference at upstream positions of highly used and focused TSSs may be leftovers of A-enrichment in those promoters during scanning evolution in addition to promoter roles beyond TSS selection per se.

Our studies highlight the strength of approaches to minimize contextual factors by isolating specific promoter attributes for study in a high-throughput fashion. Here we have employed Pol II MASTER to the DNA sequence determinants of initiation efficiency during Pol II scanning. It will be valuable to apply this systematic analysis to other promoter architectural factors determining Pol II initiation, such as UAS identity, core promoter-TSS distance, and sequence composition within the scanning region. In addition, we have found sequences downstream of TSSs contribute to TSS efficiency. Therefore, expanding the randomized region is needed to refine our understanding of sequence preference and potential longer range sequence interactions. Furthermore, applying Pol II MASTER across initiation mutants and promoter variants will reveal factor-sequence relationships and may allow initiation potential to be determined from DNA sequence and genome location alone.

2.4 Methods

2.4.1 Yeast strains, plasmids, oligonucleotides and media

Yeast strains, plasmids, and oligonucleotide sequences are described in **Appendix B** as **Appendix Table 3, 7, 11**. All oligonucleotides were obtained from IDT. Yeast strains used in this study were constructed as previously (Braberg et al., 2013; Jin and Kaplan, 2014; Kaplan et al., 2012; Qiu et al., 2020). Briefly, plasmids containing *rpb1* mutants (G1097D, E1103G, F1086S, and H1085Q) were introduced by transformation into yeast strain CKY749 containing a chromosomal deletion of *RPO21/RPB1* but with a wild type *RPB1 URA3* plasmid, which was subsequently lost by plasmid shuffling. Yeast media are following standard protocols (Amberg et al., 2005). YPD solid medium is made of yeast extract (1% w/v; BD), peptone (2% w/v; BD, 211677), bacto-agar (2% w/v; BD, 214010) and dextrose (2% w/v; VWR, VWRBK876) supplemented with adenine (0.15 mM; Sigma-Aldrich, A9126) and L-tryptophan (0.4 mM; Sigma-Aldrich T0254). Minimal media plates are synthetic complete ("SC") with amino-acids dropped out as appropriate as described in (Amberg et al., 2005) with minor alterations as described in (Kaplan et al., 2012): per standard batch formulation, adenine hemisulfate (Sigma-Aldrich, A9126) was 2 g, L-Leucine (Sigma-Aldrich, L8000) was 4 g, myo-inositol was 0.1 g, para-aminobenzoic acid (PABA) was 0.2 g.

2.4.2 Construction and transformation of plasmid libraries

A 9 nt randomized TSS region and 20 nt randomized barcodes, with 4 fixed bases inserted between every 4 nt (NNNNANNNNCNNNNTNNNNGNNNN), were separately synthesized by

IDT as oligo pools with specific randomized positions using "hand mixing" for N positions to ensure even randomization and avoid bias during machine mixing of precursors during oligo synthesis. Together with other components including the *GAL1* UAS, *SNR37* core promoter, *SNR37* TSS region ("flux detector"), *GFP* ORF, and the *CYC1* terminator, template libraries were constructed by PCR sewing and cloned into pRSII413 (a gift from Steven Haase, Addgene plasmid #35450; http://n2t.net/addgene:35450; RRID:Addgene_35450) (Chee and Haase, 2012) by ligation (**Figure 11A**). Ligation products were transformed into *Escherichia coli* TOP10F′ cells (Invitrogen) and grown on LB plates supplemented with carbenicillin (100 µg/ml) at high density. 200,000-500,000 colonies were collected from each library to maximize variant representation. Plasmid libraries were isolated from cells pellets using ZymoPURE II Plasmid Maxiprep Kit (Zymo Research, D4203) per manufacturer's instructions. Plasmid library pools were transformed into yeast strains with wildtype and mutated Pol II using chemical transformation and electroporation, respectively. For Pol II WT libraries, 500 ng plasmid pool per reaction was transformed following yeast high-efficiency transformation protocol described in (Gietz and Schiestl, 2007). For Pol II mutant libraries, 2 µg plasmid pool per reaction was electroporated into Pol II mutant strains following yeast electroporation transformation protocol described in (Benatuil et al., 2010), with 50 µg single-stranded carrier DNA added. Transformants were grown on selective SC-His plates with 2% glucose as carbon source at high density. Three biological replicates were performed for each library and on average over two million colonies were collected for each replicate. Transformants scraped from densely plated transformation plates were inoculated into fresh SC-His medium with 2% raffinose (Amresco, J392) at 0.25 x 10^7 cells/ml and grown until 0.5-0.8 x 10^7 cells/ml, as determined by cell counting. Subsequently, galactose (Amresco, 0637) was added for three hours (4% final concentration) to induce library expression.

50 ml and 5 ml culture aliquots, for RNA and DNA extraction respectively, were harvested, and then cell pellets were stored at -80 °C for further processing as described below.

2.4.3 Generation of DNA amplicon for DNA-seq

Plasmid DNA from yeast cell pellets was isolated using YeaStar Genomic DNA Kit (Zymo Research, D2002) per manufacturer's instructions. Amplicon pools containing the TSS and barcode regions were generated using plasmid DNA from *E. coli* or yeast by Micellula DNA Emulsion & Purification (ePCR) Kit (EURx/CHIMERx, 3600) per manufacturer's instructions. To minimize amplification bias, each sample was amplified in a 15-cycle ePCR reaction, purified, and subject to an additional 10-cycle scale-up ePCR reaction. To create the necessary sequence template diversity for Illumina sequencing, 18-25 bp and 1-7 bp "stuffer" sequences were added to 5′- and 3′-ends, respectively, during amplicon preparation. Amplicon pools were subject to Illumina NovaSeq 6000 (150 PE) sequencing, and on average 20 M paired-end reads were obtained from each replicate of a sample, with high reproducibility and minimal perturbation of the variant distribution with each library (**Table 3**).

Table 3 Statistics and data accession information for DNA-seq

Super _ID	Sample name	Library	*E. coli* / Yeast replicates	BioSample accession	Sequencing platform	SRA accession	GEO accession	Total paired-end reads	Assembled R1R2 merged reads by PEAR
1	AYR_Ec	AYR	*E. coli*	SAMN21846366	Illumina NovaSeq 6000 S2 150PE	SRR16089335	GSM5610210	20,778,297	20,731,982
2	BYR_Ec	BYR	*E. coli*	SAMN21846367	Illumina NovaSeq 6000 S2 150PE	SRR16089334	GSM5610211	20,367,814	20,312,301
3	ARY_Ec	ARY	*E. coli*	SAMN21846368	Illumina NovaSeq 6000 S2 150PE	SRR16089336	GSM5610212	23,909,997	23,835,147
4	AYR_PolII_WT_bioRep_1	AYR	Yeast Replicate 1	SAMN21846369	Illumina NovaSeq 6000 S2 150PE	SRR16089933	GSM5610213	22,448,161	22,398,417
5	AYR_PolII_WT_bioRep_2	AYR	Yeast Replicate 2	SAMN21846370	Illumina NovaSeq 6000 S2 150PE	SRR16089932	GSM5610214	21,907,471	21,870,559
6	AYR_PolII_WT_bioRep_3	AYR	Yeast Replicate 3	SAMN21846371	Illumina NovaSeq 6000 S2 150PE	SRR16089930	GSM5610215	17,347,170	17,323,744
7	BYR_PolII_WT_bioRep_1	BYR	Yeast Replicate 1	SAMN21846372	Illumina NovaSeq 6000 S2 150PE	SRR16089929	GSM5610216	16,977,094	16,956,572
8	BYR_PolII_WT_bioRep_2	BYR	Yeast Replicate 2	SAMN21846373	Illumina NovaSeq 6000 S2 150PE	SRR16089931	GSM5610217	20,810,116	20,767,897
9	BYR_PolII_WT_bioRep_3	BYR	Yeast Replicate 3	SAMN21846374	Illumina NovaSeq 6000 S2 150PE	SRR16089928	GSM5610218	22,939,289	22,894,586
10	ARY_PolII_WT_bioRep_1	ARY	Yeast Replicate 1	SAMN21846375	Illumina NovaSeq 6000 S2 150PE	SRR16089927	GSM5610219	24,932,189	24,904,059
11	ARY_PolII_WT_bioRep_2	ARY	Yeast Replicate 2	SAMN21846376	Illumina NovaSeq 6000 S2 150PE	SRR16089926	GSM5610220	19,133,313	19,099,282
12	ARY_PolII_WT_bioRep_3	ARY	Yeast Replicate 3	SAMN21846377	Illumina NovaSeq 6000 S2 150PE	SRR16089925	GSM5610221	20,111,542	20,063,548

2.4.4 Sample preparation for TSS-seq

Total RNA was extracted by a phenol-chloroform method (Schmitt et al., 1990), followed by RNA purification (RNeasy Mini kit, QIAGEN, 74104) with on-column DNase digestion (RNase-Free DNase Set, QIAGEN, 79254) to remove DNA. TSS-seq was done using procedures described in (Vvedenskaya et al., 2015a). To prepare RNAs for the cDNA library construction, samples were sequentially treated with Terminator 5′-Phosphate-Dependent Exonuclease (Lucigen), Quick CIP (calf-intestine alkaline phosphatase, NEB), and Cap-Clip™ Acid Pyrophosphatase (CellScript) to remove 5′ monophosphate RNA and convert 5′ triphosphate or capped RNAs to 5′ monophosphate RNAs. Next, RNA prepared with enzymatic treatments was ligated to the 5′-adapter (s1206-N15, 5′-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNNNNNNNNNNN-3′) that contains Illumina adapter sequence and a 15 nt randomized 3′-end to reduce ligation bias and serve as a Unique Molecular Identifier (UMI). Next, cDNA was constructed by reverse transcription using RT primer CKO2191-s128A (5′-CCTTGGCACCCGAGAATTCCAAGTGAATAATTCTTCACCTTTA-3′) followed by emulsion PCR amplification for 20-22 cycles using Illumina PCR primers (RP1 and RPI3-30). Final DNA was gel size selected for 180-250 bp lengths and sequenced by Illumina NextSeq 500 (150 SE) or NovaSeq 6000 (200 SE) using custom primer s1115 (5′-CTACACGTTCAGAGTTCTACAGTCCGACGATC-3′) to avoid potentially confounding effects of misannealing of the default pooled Illumina sequencing primers to the two randomized sequence regions (**Table 4**).

Table 4 Statistics and data accession information for TSS-seq

Super_ID	Sample name	BioSample #	Batch-1 sequencing platform	Batch-1 SRA #	Batch-1 GEO #	Total reads of batch-1	Batch-2 sequencing platform	Batch-2 SRA #	Batch-2 GEO #	Total reads of batch-2	Total reads of 2 batches
4	AYR_PolII_WT_bioRep_1	SAMN21846369	Illumina NextSeq 500 150SE	SRR16096607	GSM5610222	31,214,370					31,214,370
5	AYR_PolII_WT_bioRep_2	SAMN21846370	Illumina NextSeq 500 150SE	SRR16096606	GSM5610223	52,215,117					52,215,117
6	AYR_PolII_WT_bioRep_3	SAMN21846371	Illumina NextSeq 500 150SE	SRR16096595	GSM5610224	52,276,068					52,276,068
7	BYR_PolII_WT_bioRep_1	SAMN21846372	Illumina NextSeq 500 150SE	SRR16096587	GSM5610225	46,913,687	Illumina NovaSeq 6000 S1 200SE	SRR16097928	GSM5610267	30,227,969	77,141,656
8	BYR_PolII_WT_bioRep_2	SAMN21846373	Illumina NextSeq 500 150SE	SRR16096586	GSM5610226	50,842,225	Illumina NovaSeq 6000 S1 200SE	SRR16097927	GSM5610268	24,592,663	75,434,888
9	BYR_PolII_WT_bioRep_3	SAMN21846374	Illumina NextSeq 500 150SE	SRR16096585	GSM5610227	64,556,129	Illumina NovaSeq 6000 S1 200SE	SRR16097916	GSM5610269	37,189,489	101,745,618
10	ARY_PolII_WT_bioRep_1	SAMN21846375	Illumina NextSeq 500 150SE	SRR16096584	GSM5610228	49,546,395					49,546,395
11	ARY_PolII_WT_bioRep_2	SAMN21846376	Illumina NextSeq 500 150SE	SRR16096583	GSM5610229	51,548,429					51,548,429
12	ARY_PolII_WT_bioRep_3	SAMN21846377	Illumina NextSeq 500 150SE	SRR16096582	GSM5610230	46,993,934					46,993,934
13	AYR_PolII_E1103G_bioRep_1	SAMN21846378	Illumina NextSeq 500 150SE	SRR16096581	GSM5610231	58,973,392	Illumina NovaSeq 6000 S1 200SE	SRR16097908	GSM5610270	42,090,759	101,064,151
14	AYR_PolII_E1103G_bioRep_2	SAMN21846379	Illumina NextSeq 500 150SE	SRR16096605	GSM5610232	48,112,709	Illumina NovaSeq 6000 S1 200SE	SRR16097907	GSM5610271	42,157,919	90,270,628

Table 4 (continued).

15	AYR_PolII_E1103G_bioRep_3	SAMN21846380	Illumina NextSeq 500 150SE	SRR16096604	GSM5610233	76,198,791	Illumina NovaSeq 6000 S1 200SE	SRR16097906	GSM5610272	47,196,236	123,395,027
16	BYR_PolII_E1103G_bioRep_1	SAMN21846381	Illumina NextSeq 500 150SE	SRR16096603	GSM5610234	60,866,054	Illumina NovaSeq 6000 S1 200SE	SRR16097905	GSM5610273	91,555,567	152,421,621
17	BYR_PolII_E1103G_bioRep_2	SAMN21846382	Illumina NextSeq 500 150SE	SRR16096602	GSM5610235	52,700,626	Illumina NovaSeq 6000 S1 200SE	SRR16097904	GSM5610274	121,543,313	174,243,939
18	BYR_PolII_E1103G_bioRep_3	SAMN21846383	Illumina NextSeq 500 150SE	SRR16096601	GSM5610236	66,426,468	Illumina NovaSeq 6000 S1 200SE	SRR16097903	GSM5610275	102,771,964	169,198,432
19	ARY_PolII_E1103G_bioRep_1	SAMN21846384	Illumina NextSeq 500 150SE	SRR16096600	GSM5610237	71,153,395	Illumina NovaSeq 6000 S1 200SE	SRR16097902	GSM5610276	19,019,139	90,172,534
20	ARY_PolII_E1103G_bioRep_2	SAMN21846385	Illumina NextSeq 500 150SE	SRR16096599	GSM5610238	47,447,124	Illumina NovaSeq 6000 S1 200SE	SRR16097926	GSM5610277	25,961,742	73,408,866
21	ARY_PolII_E1103G_bioRep_3	SAMN21846386	Illumina NextSeq 500 150SE	SRR16096598	GSM5610239	54,000,114	Illumina NovaSeq 6000 S1 200SE	SRR16097925	GSM5610278	25,534,089	79,534,203
22	AYR_PolII_F1086S_bioRep_1	SAMN21846387	Illumina NextSeq 500 150SE	SRR16096597	GSM5610240	56,977,919					56,977,919
23	AYR_PolII_F1086S_bioRep_2	SAMN21846388	Illumina NextSeq 500 150SE	SRR16096596	GSM5610241	49,516,432					49,516,432
24	AYR_PolII_F1086S_bioRep_3	SAMN21846389	Illumina NextSeq 500 150SE	SRR16096594	GSM5610242	48,864,974					48,864,974
25	BYR_PolII_F1086S_bioRep_1	SAMN21846390	Illumina NextSeq 500 150SE	SRR16096593	GSM5610243	52,966,627	Illumina NovaSeq 6000 S1 200SE	SRR16097924	GSM5610279	107,498,631	160,465,258
26	BYR_PolII_F1086S_bioRep_2	SAMN21846391	Illumina NextSeq 500 150SE	SRR16096592	GSM5610244	56,137,472	Illumina NovaSeq 6000 S1 200SE	SRR16097923	GSM5610280	97,534,318	153,671,790

Table 4 (continued).

27	BYR_PolII_F1086S_bioRep_3	SAMN21846392	Illumina NextSeq 500 150SE	SRR16096591	GSM5610245	44,960,940	Illumina NovaSeq 6000 S1 200SE	SRR16097922	GSM5610281	106,042,643	151,003,583
28	ARY_PolII_F1086S_bioRep_1	SAMN21846393	Illumina NextSeq 500 150SE	SRR16096590	GSM5610246	38,195,316					38,195,316
29	ARY_PolII_F1086S_bioRep_2	SAMN21846394	Illumina NextSeq 500 150SE	SRR16096589	GSM5610247	44,913,456					44,913,456
30	ARY_PolII_F1086S_bioRep_3	SAMN21846395	Illumina NextSeq 500 150SE	SRR16096588	GSM5610248	47,080,210					47,080,210
33	AYR_PolII_G1097D_bioRep_1	SAMN21846396	Illumina NextSeq 500 150SE	SRR16097356	GSM5610249	22,496,621	Illumina NovaSeq 6000 S1 200SE	SRR16097921	GSM5610282	80,975,069	103,471,690
34	AYR_PolII_G1097D_bioRep_2	SAMN21846397	Illumina NextSeq 500 150SE	SRR16097355	GSM5610250	21,295,790	Illumina NovaSeq 6000 S1 200SE	SRR16097920	GSM5610283	91,359,713	112,655,503
35	AYR_PolII_G1097D_bioRep_3	SAMN21846398	Illumina NextSeq 500 150SE	SRR16097346	GSM5610251	21,420,780	Illumina NovaSeq 6000 S1 200SE	SRR16097919	GSM5610284	76,106,035	97,526,815
36	BYR_PolII_G1097D_bioRep_2	SAMN21846399	Illumina NextSeq 500 150SE	SRR16097345	GSM5610252	72,748,369	Illumina NovaSeq 6000 S1 200SE	SRR16097918	GSM5610285	248,760,587	321,508,956
37	BYR_PolII_G1097D_bioRep_4	SAMN21846400	Illumina NextSeq 500 150SE	SRR16097344	GSM5610253	62,131,585	Illumina NovaSeq 6000 S1 200SE	SRR16097917	GSM5610286	246,721,585	308,853,170
38	BYR_PolII_G1097D_bioRep_5	SAMN21846401	Illumina NextSeq 500 150SE	SRR16097343	GSM5610254	70,373,887	Illumina NovaSeq 6000 S1 200SE	SRR16097915	GSM5610287	194,980,132	265,354,019
41	ARY_PolII_G1097D_bioRep_3	SAMN21846402	Illumina NextSeq 500 150SE	SRR16097342	GSM5610255	20,837,569	Illumina NovaSeq 6000 S1 200SE	SRR16097914	GSM5610288	20,401,590	41,239,159
42	ARY_PolII_G1097D_bioRep_5	SAMN21846403	Illumina NextSeq 500 150SE	SRR16097341	GSM5610256	26,252,292	Illumina NovaSeq 6000 S1 200SE	SRR16097913	GSM5610289	20,823,983	47,076,275

Table 4 (continued).

43	ARY_PolII_G1097D_bioRep_6	SAMN21846404	Illumina NextSeq 500 150SE	SRR16097340	GSM5610257	29,254,270	Illumina NovaSeq 6000 S1 200SE	SRR16097912	GSM5610290	23,023,279	52,277,549
44	AYR_PolII_H1085Q_bioRep_1	SAMN21846405	Illumina NextSeq 500 150SE	SRR16097339	GSM5610258	28,258,455					28,258,455
45	AYR_PolII_H1085Q_bioRep_2	SAMN21846406	Illumina NextSeq 500 150SE	SRR16097354	GSM5610259	25,727,329					25,727,329
46	AYR_PolII_H1085Q_bioRep_3	SAMN21846407	Illumina NextSeq 500 150SE	SRR16097353	GSM5610260	23,359,504					23,359,504
47	BYR_PolII_H1085Q_bioRep_1	SAMN21846408	Illumina NextSeq 500 150SE	SRR16097352	GSM5610261	80,275,208	Illumina NovaSeq 6000 S1 200SE	SRR16097911	GSM5610291	118,098,417	198,373,625
48	BYR_PolII_H1085Q_bioRep_2	SAMN21846409	Illumina NextSeq 500 150SE	SRR16097351	GSM5610262	86,833,431	Illumina NovaSeq 6000 S1 200SE	SRR16097910	GSM5610292	102,501,325	189,334,756
49	BYR_PolII_H1085Q_bioRep_3	SAMN21846410	Illumina NextSeq 500 150SE	SRR16097350	GSM5610263	85,381,271	Illumina NovaSeq 6000 S1 200SE	SRR16097909	GSM5610293	122,222,426	207,603,697
50	ARY_PolII_H1085Q_bioRep_1	SAMN21846411	Illumina NextSeq 500 150SE	SRR16097349	GSM5610264	28,654,160					28,654,160
51	ARY_PolII_H1085Q_bioRep_2	SAMN21846412	Illumina NextSeq 500 150SE	SRR16097348	GSM5610265	27,235,701					27,235,701
52	ARY_PolII_H1085Q_bioRep_3	SAMN21846413	Illumina NextSeq 500 150SE	SRR16097347	GSM5610266	31,376,531					31,376,531

2.4.5 Primer extension assay

Primer extension assays were performed on the same batch of total RNA extracted for TSS-seq as described in (Ranish and Hahn, 1991) with modifications described in (Kaplan et al., 2012). For each reaction, 30 µg total RNA was used. An RNA sample without library transformed was used as "no GFP" control. A sample containing the same amount of nuclease-free water was used as "no RNA" control. A primer (CKO2191) complementary to the 6th to 27th bases of *GFP* ORF, which is the same annealing region for reverse transcription of TSS-seq sample preparation, was labeled with ^{32}P γ-ATP (PerkinElmer, BLU502Z250UC) and T4 polynucleotide kinase (Thermo Scientific, EK0031). M-MuLV Reverse Transcriptase (NEB, M0253L), RNase inhibitor (NEB, M0307L), dNTPs (GE), and DTT were added to mix of RNA and labelled primer for reverse transcription reaction. Before loading to sequencing gel, RNase A (Thermo Scientific, EN0531) was added to remove RNA. The products were analyzed by 8% acrylamide/bis-acrylamide (19:1 ratio, Bio-Rad, 1610145) gel containing 1x TBE and 7M urea. Primer extension gel was visualized by Molecular Imager PharosFX™ Plus System (Bio-Rad) and quantified by Image Lab (5.2).

2.4.6 Computational analyses

Data and statistical analyses were performed in Python (3.8.5) and R (4.0.0) environments. Additional packages usages are reported throughout the methods description. Source code is provided at https://github.com/Kaplan-Lab-Pitt/PolII_MASTER-TSS_sequence. Raw sequencing data have been deposited on the NCBI SRA (Sequence Read Archive) under the BioProject accession number PRJNA766624. Processed data have been deposited on the GEO (Gene

Expression Omnibus) under the accession number GSE185290. Visualizations were compiled in Adobe Illustrator 2021.

2.4.6.1 DNA-seq analysis

High-throughput sequencing of template DNA amplicon was used to assign each 9 nt randomized TSS sequence to a corresponding 24 nt barcode. First, paired-end reads were merged using PEAR (0.9.11) (Zhang et al., 2014). Next, we considered only those reads that contained a perfect match to three sequence regions common to all variants: 27 nt sequence upstream of the TSS region, 24 nt sequence between the TSS region and the barcode region, and 27 nt sequence downstream of the barcode region (5′-TTCAAATTTTTCTTTTGATTTTTTTTCNNNNNNNNNACATTTTCAAAAGGCTAACATCAGNNNNANNNNCNNNNTNNNNGNNNNATGTCTAAAGGTGAAGAATTATTCACT-3′, randomized TSS and barcode regions are underlined). From these reads, 9 nt TSS region and 24 nt barcode were extracted, followed by individual error correction using UMI-tools (1.0.0) (Smith et al., 2017). Next, for barcodes linking to multiple TSS variants, only barcodes for which ≥ 90% of the sequencing reads containing a specified barcode also contained a shared, exact 9 nt TSS region were kept. To generate a master pool of TSS-barcode linkages for all TSS-seq samples, for each library (“AYR”, “BYR”, “ARY”), TSS-barcode linkages that existed in at least two out of four samples (one *E. coli* sample plus three WT yeast replicates) and in which ≥ 5 reads existed were kept and pooled. Two types of processed data are available in the GEO database, with accession numbers listed in (**Table 3**): tables containing TSS-barcode linkages and corresponding DNA-seq read counts for each sample, tables of the master pool containing kept TSS-barcode linkages and corresponding DNA-seq read count in all related samples.

2.4.6.2 TSS-seq analysis for libraries

High-throughput sequencing of RNA samples was used to link RNA products to barcodes, therefore assigning TSS usage to corresponding DNA templates. We considered only those reads that contained a perfect match to a 27 nt sequence region downstream of the barcode region, as well as expected length of 5′-end: 5′-[15 nt 5′-UMI]-[>1 nt upstream of the barcode region, designated as "RNA 5′-end"]-[24 nt barcode]-[the first 27 nt of *GFP* ORF, ATGTCTAAAGGTGAAGAATTATTCACT]-3′. Next, 15 nt 5′-UMIs, "RNA 5′-end" with varying length, and 24 nt barcode were extracted and individually corrected by UMI-tools (1.0.0). Deduplication was performed based on 5′-UMIs, meaning reads contained a shared UMI-"RNA 5′-end"-barcode linkage were counted as one deduplicated read for further analysis. Next, the identity of the 24 nt barcode was used to determine the template sequences of the randomized TSS region. Then, reads with "RNA 5′-end" sequence perfectly matched to corresponding template sequence were kept and used for downstream analysis. A TSS-seq count table containing TSS usage distribution of each TSS promoter variant was generated. In the count table, each row represents one TSS promoter variant, and each column represents one position between positions -68 to +25 relative to "designed" +1 TSS. The number in each cell represents TSS-seq reads generated from a particular position, with perfectly match to DNA template. After investigating reproducibility, count tables generated from three biological replicates were merged into one by aggregating read counts at each position. Promoter variants with ≥ 5 TSS-seq reads in each replicate and whose Coefficient of Variation (CV) of TSS-seq reads is less than 0.5 were kept. Two types of processed data are available in the GEO database, with accession numbers listed in (**Table 4**): tables containing "RNA 5′-end"-barcode linkages and corresponding deduplicated TSS-

seq read counts for individual sample, TSS-seq count tables for individual samples and after aggregating replicates.

TSS efficiency for each position was calculated by dividing reads count at a particular position by the reads at or downstream of this TSS. TSS positions with ≥ 20% efficiency but with ≤ 5 reads left for this and following positions were filtered out, as well as their downstream positions.

2.4.6.3 Sequence preference analysis

For sequence preference at each position, all TSS variants were subgrouped based on the bases at a particular position. TSS efficiencies of TSS variants were visualized as scatter plots using GraphPad Prism 9. Kruskal-Wallis with Dunn's test was performed to test sequence preference in GraphPad Prism 9. Next, TSS efficiency medians of each subgroup were calculated and centered to calculate "relative efficiency" at each position. The relative efficiencies were visualized as sequence logos using Logomaker (0.8) (Tareen and Kinney, 2020). In motif enrichment analysis, surrounding sequences relative to examined TSSs were extracted and visualized as sequence logos using WebLogo 3 (Crooks et al., 2004). Heatmaps, scatter plots, and density plots for comparing Pol II WT and mutants were generated by Morpheus (https://software.broadinstitute.org/morpheus) or ggplot2 (3.3.3) R package.

2.4.6.4 Interaction analysis

The interaction between positions is defined as different bases existing at one position resulting in different sequence preferences at another position. For any two positions, all TSS variants were subgrouped based on bases at both positions. Median values of TSS efficiency distribution of each subgroup were calculated and centered twice to calculate "centered relative

efficiency". The centered relative efficiencies were visualized as heatmaps using Seaborn (0.11.0). Interactions related to positions between -11 to -9 were calculated using datasets of designed +4 TSSs deriving from "AYR", "BYR", and "ARY" libraries. Other interactions were calculated using datasets of designed +1 TSSs deriving from "AYR" and "BYR" libraries.

2.4.6.5 TSS-seq analysis for genomic TSSs.

Genomic TSS-seq datasets are from our lab's previous study (Zhao et al., 2021). Quality control, read trimming, and mapping were performed as described in (Qiu et al., 2020; Zhao et al., 2021) to generate a TSS count table that contains TSS-seq reads at each position within known promoter windows ("median" TSS, 250 nt upstream and 150 nt downstream from median TSS position). TSS efficiency calculation and subsequent sequence preference analyses were performed as that for Pol II MASTER libraries.

2.4.6.6 Prediction of TSS efficiency

To prepare datasets for modeling, positions of designed -8 to +2 and +4 TSSs of each promoter variant that have valid TSS efficiency were compiled as TSS sequence variants. For each TSS variant, sequences at positions -11 to +9 relative to TSS, together with corresponding TSS efficiency, were extracted. 80% of the dataset were randomly partitioned as training set and the rest 20% as testing set. To select robust features, a forward stepwise strategy with a 5-fold Cross-Validation (CV) was employed in two major stages, for additive terms and for interactions. Starting with no variable in the model, logistic regression models with one additional variable (the sequence at a particular position) were trained to predict TSS efficiency on the training set by train() of caret (6.0.86) R package (Kuhn, 2008), with a 5-fold CV. The R^2, representing the proportion of variance explained, was calculated to indicate the performance of each model. The

variable that provides the highest increased R^2 for model was added into the model for the next round of variable selection. This process was repeated until the increased R^2 is less than 0.01. After identifying the most influential additive variables, the same process was repeated for investigating robust interactions between selected additive variables. Next, a final model with selected robust features, including additive variables and interactions, was constructed on the entire training set using glm() and investigated on the testing set. Comparison between predicted and measured efficiencies was visualized as scatter plots using LSD (4.1.0) R package. Model parameters were extracted and used to further calculation. Visualizations were done in Logomaker (0.8) and Seaborn (0.11.0) in Python. Principal Component Analysis (PCA) was performed using prcomp() in R.

3.0 Determination of the role of Tfb3 in Pol II transcription initiation

3.1 Introduction

Tfb3 is one of eleven subunits of the conserved eukaryotic General Transcription Factor (GTF) TFIIH in *Saccharomyces cerevisiae*. TFIIH functions in multiple fundamental processes of RNA Pol II transcription such as promoter opening and Pol II CTD phosphorylation (Jona et al., 2002; Schilbach et al., 2021; Schilbach et al., 2017; van Eeuwen et al., 2021), Nucleotide Excision Repair (NER) of DNA damage (Feaver et al., 2000), and cullin activation of some E3 ligases (Rabut et al., 2011). Structural studies indicate that, during transcription initiation, the TFIIH component Tfb3 links TFIIH to the rest of the Pre-initiation Complex (PIC) via its interactions with Pol II stalk subcomplex Rpb4-Rpb7, the GTF TFIIE, and other TFIIH components for promoter scanning (Luo et al., 2015; Schilbach et al., 2021; Schilbach et al., 2017). Pol II, TFIIE, and TFIIH have important roles in the TSS selection process indicated by previous genetic and biochemical studies. However, how Tfb3 or its proposed interactions participate in TSS selection with other GTFs has not been directly examined.

Pol II initiation in yeast proceeds by promoter scanning, where the PIC assembles upstream of a promoter and scans downstream to select TSSs. Studies on Pol II and GTF mutant effects on TSS distribution further proposed that two functional networks contribute to TSS selection during promoter scanning. The "**efficiency**" network, including Pol II, TFIIB, and TFIIF, contributes to TSS distribution via regulating initiation efficiency, namely the probability of initiating transcription from a particular TSS. The "**processivity**" network, including TFIIH subunit Ssl2, Sub1, and potentially TFIIF, contributes to TSS distribution via TFIIH processivity, namely the

probability a particular TSS can be reached by Pol II. Tfb3 has been proposed to bridge TFIIH to the rest of the PIC via physical interactions with Pol II stalk subcomplex Rpb4-Rpb7, TFIIE, and TFIIH subunit Rad3 during promoter scanning (Schilbach et al., 2021; Schilbach et al., 2017). These physical connections suggested that Tfb3 has the potential to also functionally connect Pol II and TFIIH and the two networks in which they are involved. Recent mammalian PIC structures suggested that TFIIH detaches from the rest of the PIC during initiation, as MAT1, the Tfb3 homolog in humans, has been observed to be mobile and loses its connection with the Pol II stack complex (Aibara et al., 2021). This conformational change during initiation may also regulate TFIIH activity during scanning in yeast, as the PIC structures and connections between TFIIH and the rest of the PIC are largely conserved from humans to yeast (Aibara et al., 2021; Schilbach et al., 2021; Schilbach et al., 2017). However, the extent of these contacts is different. The more extensive connections in yeast, compared to that in higher eukaryotes, potentially delay the detachment of TFIIH from the rest of PIC and therefore support the longer scanning distance, namely the usage of further downstream TSSs. This hypothesis is consistent with a previous observation that omission of TFIIK (containing Tfb3, Kin28, and Ccl1), predicted to disrupt those connections, shifted TSS usage to upstream to similar position as that in higher eukaryotes (Murakami et al., 2015). However, the proposed functions of those contacts within the PIC via Tfb3 have not been directly tested yet. Moreover, as a subunit of TFIIH, it is important to study how Tfb3 functionally interacts with Ssl2 and its processivity. Therefore, a comprehensive study of Tfb3 function in TSS selection and its functional interaction with other initiation factors is needed to understand the scanning mechanism.

Here, I performed a genetic screen for *tfb3* mutants with transcription-related phenotypes to understand Tfb3 function in Pol II initiation in *S. cerevisiae*. I identified two classes of *tfb3*

mutants conferring distinct transcriptional phenotypes *in vivo* and resulting in polar shifts in TSS usage. One class of mutants conferred defects in *IMD2* induction, resulting in sensitivity to the drug MPA. We and others have established this phenotype as predictive of shifting TSS usage upstream globally (Jenks et al., 2008; Kaplan et al., 2012; Kuehner and Brow, 2008; Qiu et al., 2020; Zhao et al., 2021). Another class of mutants produced functional *HIS3* mRNA using an *imd2Δ::HIS3* reporter, which is predictive of a shifting of TSS usage downstream. Both classes were validated by examination of TSS usage at *ADH1*. Further, I quantitatively detected that *tfb3* mutants shifted TSS distribution upstream via increasing usage for both upstream minor and major TSSs, which is a pattern combining features of "efficiency" upstream shifters (Pol II GOF alleles) and "processivity" upstream shifters (Ssl2 LOF alleles), suggesting Tfb3 might be involved in both efficiency and processivity networks that control transcription initiation. Moreover, I investigated genetic interactions of *tfb3* alleles with *ssl2* alleles or *sub1*Δ. The observed primary additive interactions supported Tfb3 functions in regulating TFIIH′s processivity.

3.2 Results

3.2.1 Transcription-related phenotypes

To identify potential *tfb3* mutants that have effects on Pol II initiation, four transcription-related phenotypes that have been demonstrated to be predictive of specific initiation defects have been used: MPA^S, His^+, GAL^R, and Spt^- (**Figure 22**) (Zhao et al., 2021). Each of them is discussed below.

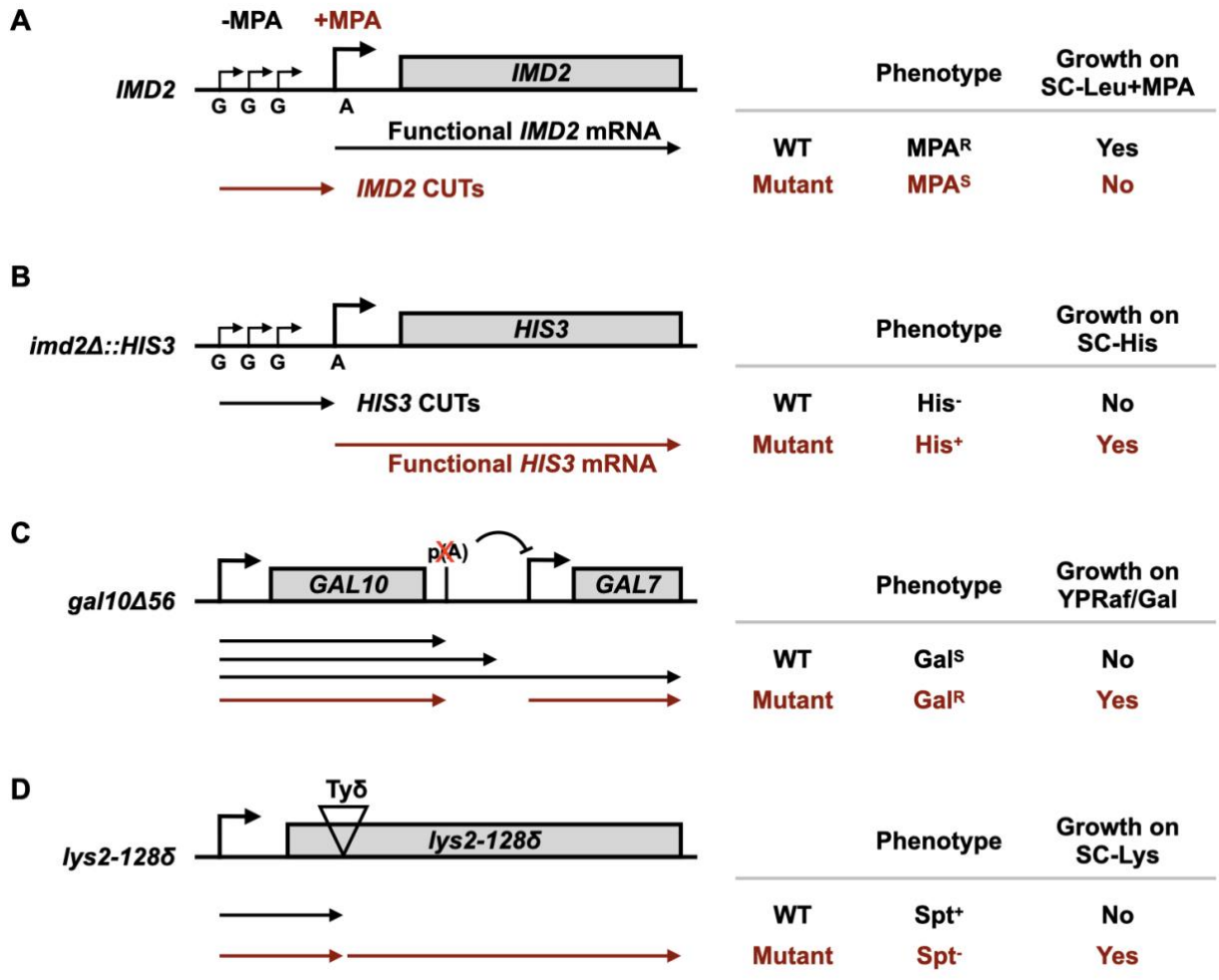


Figure 22 Transcriptional phenotypes used in *tfb3* mutant screening

Schematic representation of four transcriptional phenotypes used in *tfb3* mutant screening. **(A)** MPA^S (Mycophenolic Acid sensitivity). The drug MPA inhibits yeast inosine monophosphate dehydrogenase (IMPDH) activities encoded by *IMD3* and *IMD4*; however, the GTP starvation induces *IMD2* expression by shifting TSS usage from upstream "G" TSSs (producing non-functional CUTs) to a downstream "A" TSS (producing functional *IMD2* transcript) in WT. Mutants that are unable to shift the usage to the downstream "A" TSS show MPA^S. **(B)** His$^+$. The *IMD2* ORF was replaced with the *HIS3* ORF to construct the *imd2Δ::HIS3* reporter. In the absence of MPA, upstream TSSs of *IMD2* promoter producing non-functional transcripts are used, therefore WT cells cannot grow on medium lacking histidine (SC-His). Mutants that are able to use the downstream TSS can grow and therefore show His$^+$ phenotype. **(C)** Gal^R (Galactose Resistance in *gal10Δ56*). Deletion of the major *GAL10* poly(A) (*gal10Δ56*) causes Pol II readthrough and therefore interferes with *GAL7* initiation, resulting in galactose sensitivity for WT cells. Mutants that can rescue *GAL7*

transcription via facilitating either *GAL10* termination or *GAL7* initiation show resistance to galactose (Gal^R). **(D)** Spt^- (Suppressor of Ty). Insertion of a Ty retrotransposon into the *LYS2* gene (*lys2-128δ*) leads to no functional *LYS2* mRNA being produced. Therefore WT cells are unable to grow on medium lacking lysine (SC-Lys). Mutants that can activate the initiation from a promoter inside of *lys2-128δ* rescue *LYS2* expression and therefore can grow (Spt^-).

(1) MPA^S (Mycophenolic Acid sensitivity) (**Figure 22A**) (Jenks et al., 2008; Kuehner and Brow, 2008). MPA^S phenotype can be used as a proxy for potential upstream TSS shifts because it correlates with initiation defects at the *IMD2* gene, or more specifically, defects in using a downstream TSS required for a functional *IMD2* transcript. The drug Mycophenolic Acid (MPA) inhibits yeast inosine monophosphate dehydrogenase (IMPDH) activities encoded by two MPA-sensitive IMPDH paralogs (*IMD3* and *IMD4*) and therefore reduces GTP levels. However, WT cells are resistant to MPA because an MPA-resistant IMPDH paralog, *IMD2*, is induced by shifting TSS usage from upstream "G" TSSs (producing non-functional CUTs) to a downstream "A" TSS (producing a functional *IMD2* transcript). Therefore, transcription mutants with defects on *IMD2* expression are sensitive to MPA, designated as MPA^S mutants. In other words, cells that cannot grow on MPA-containing plates are MPA^S mutants. Importantly, MPA^S mutants are predicted to cause an upstream shift of TSS distribution, at least at the *IMD2* promoter. Previously tested mutants in Pol II (such as G1097D, L1101S, and E1103G), Ssl2 (such as N230D, D522V, and Y750*), and Tfg2 (*tfb2Δ146-180*) that confer MPA^S phenotype have been demonstrated to shift TSS distribution upstream genome-wide, as detected by genomic TSS-seq (Kaplan et al., 2012; Qiu et al., 2020; Zhao et al., 2021). Given that MPA^S mutants with global defects showed defects at *ADH1*, investigation of TSS usage at *ADH1* by Primer Extension (PE) is used as a proxy for potential global defects as it is a gene unrelated to *IMD2*.

(2) His$^+$ (**Figure 22B**) (Malik et al., 2017). His$^+$ phenotype, detected by the *imd2*Δ*::HIS3* reporter, is used as a proxy for potential TSS downstream shifting effect because it detects constitutive expression of *IMD2* promoter without induction of MPA, namely, constitutive usage of the *IMD2* downstream TSS. The *imd2*Δ*::HIS3* reporter was constructed by replacing the *IMD2* ORF with the *HIS3* ORF. As noticed previously, without MPA as inducer, upstream TSSs of *IMD2* promoter producing non-functional transcripts are used, therefore WT cells with the *imd2*Δ*::HIS3* reporter cannot grow on medium lacking histidine (SC-His) because no functional *HIS3* mRNA is produced. Transcription mutants that cause downstream TSS shifting allow functional *HIS3* mRNA to be produced and therefore can grow on SC-His plate, designated as His$^+$ mutants. Previous tested mutants in Pol II (such as F1086S, H1085Y, and H1085Q), Ssl2 (such as L225P, N230I, and R636C), TFIIB (*sua7-1*), and Sub1 (*sub1*Δ) conferring His$^+$ phenotype have been demonstrated to shift TSS distribution downstream at the *ADH1* promoter by PE and genome-wide by genomic TSS-seq (Kaplan et al., 2012; Qiu et al., 2020; Zhao et al., 2021), supporting His$^+$ phenotype as a good indicator for TSS downstream shifters.

(3) Gal^R (Galactose Resistance in *gal10*Δ*56*) (**Figure 22C**) (Greger and Proudfoot, 1998; Kaplan et al., 2005). Deletion of the major *GAL10* poly(A) (*gal10*Δ*56*) causes Pol II readthrough and therefore interferes with *GAL7* initiation (which is transcribed immediately adjacent and downstream of *GAL10*), resulting in galactose sensitivity. Transcription mutants that cause either *GAL10* termination or *GAL7* initiation and therefore rescue *GAL7* transcription become resistant to galactose, designated as Gal^R mutants.

(4) Spt$^-$ (Suppressor of Ty) (**Figure 22D**) (Simchen et al., 1984). The *lys2-128δ* reporter contains an insertion of a Ty retrotransposon into the *LYS2* gene. This insertion leads to no functional *LYS2* mRNA produced and therefore results in a lysine auxotrophy, namely an inability

to grow on medium lacking lysine (SC-Lys). Transcription mutants that can activate the initiation from a promoter inside of *lys2-128δ* and therefore rescue *LYS2* expression can suppress *lys2-128δ* and are designated as Spt$^-$ mutants.

In addition, conditional phenotypes including Temperature-sensitive (Ts, 37°C) and Cold-sensitive (Cs, 16°C) phenotypes have also been included in the *TFB3* screen to identify mutations that have effects on protein structure, function, and assembly.

3.2.2 *tfb3* mutant screen with gap-repair strategy and transcription-related phenotypes

The *tfb3* mutant screen was performed by a gap-repair strategy. The DNA insert fragment was generated by PCR-based random mutagenesis with mutagenic PCR using primers designed to flank 500 bp 5′ UTR, ORF, and 100 bp 3′ UTR of the *TFB3* gene. Linearized *LEU2* plasmid vector was prepared by digestion. Both fragments were transformed into yeast strains with *tfb3*Δ, *gal10*Δ*56*, *lys2-128∂*, and either *IMD2* or *imd2*Δ*::HIS3* in the genome and *TFB3* WT on a *URA3* plasmid. Transformants were grown on SC-Leu plates for three days to select cells containing gap-repaired *LEU2* plasmids. Subsequently, colonies were replicated onto plates containing 5FOA to kick out the *TFB3* WT *URA3* plasmid and therefore allow phenotypes of *tfb3* mutants to be investigated. After four days of growth, 5FOA-containing plates were replicated onto phenotyping plates, including SC-Leu+20 μg/ml MPA plate specifically for *IMD2* background, SC-His+1 mM 3AT plate for *imd2*Δ*::HIS3* background, and SC-Leu, SC-Lys, YPD (at 30°C, 37°C, 16°C), YPRaf, and YPRaf/Gal plates for both backgrounds. Phenotypes were recorded over seven days of subsequent growth.

In total, about 21,000 colonies were screened and 484 colonies showed initial transcription-related (MPA^S, His^+, and Gal^R) and/or conditional (Ts and Cs) phenotypes. After re-phenotyping, 155 colonies with relatively strong phenotypes were selected as candidates for further analysis. Plasmid DNA extracted from yeast cells was amplified in *E. coli* and then sent for Sanger sequencing to identify mutations. Among the 155 sequenced candidates, 17 candidates did not contain any nucleotide mutation within *TFB3*, while all other candidates contained 1-5 nucleotide mutations. 65 candidates contained single Amino Acid (AA) substitutions, with some mutations identified multiple times. As a result, 43 candidates containing a unique single amino acid substitution were selected for further analysis. To determine plasmid linkage of observed phenotypes, plasmids of 43 unique candidates were re-transformed into *TFB3* shuffle strains, followed by plasmid shuffle via patch and spot assays as described. These experiments resulted in 31 *tfb3* single mutants with verified phenotypes identified for further analysis. Details of these 31 *tfb3* singles are shown in **Table 5**, with spots assay results described and shown in the next section.

Table 5 *tfb3* mutants with single amino acid substitutions

***tfb3* allele #**	**Substitution position**	**Amino acid change**	**Nucleotide change**	**Confirmed phenotype(s)**
tfb3-118	7	E7K	GAG->AAG	His^+
tfb3-119	12	M12K	ATG->AAG	His^+
tfb3-84	13	C13S	TGT->AGT	MPA^S/Gal^R/Ts
tfb3-131	15	I15N	ATC->AAC	His^+
tfb3-59	16	C16R	TGT->CGT	MPA^S/Gal^R/Ts/Cs
tfb3-62	17	K17E	AAG->GAG	MPA^S/Ts
tfb3-156	17	K17R	AAG->AGG	His^+
tfb3-82	21	Y21H	TAC->CAC	MPA^S
tfb3-55	22	L22P	CTT->CCT	MPA^S/Gal^R/Ts/Cs
tfb3-60	23	S23P	TCT->CCT	MPA^S/Gal^R/Ts/Cs
tfb3-42	28	F28L	TTT->CTT	His^+
tfb3-128	38	I38F	ATC->TTC	His^+
tfb3-53	39	C39R	TGT->CGT	MPA^S/Ts/Cs
tfb3-151	44	D44G	GAT->GGT	His^+
tfb3-72	46	I46T	ATA->ACA	His^+
tfb3-19	56	Y56C	TAT->TGT	His^+
tfb3-130	56	Y56F	TAT->TTT	His^+
tfb3-43	62	I62A	ATT->GCT	His^+
tfb3-142	62	I62T	ATT->ACT	His^+
tfb3-86	64	R64K	AGA->AAA	MPA^S
tfb3-32	66	N66D	AAT->GAT	MPA^S
tfb3-106	68	F68L	TTC->CTC	MPA^S
tfb3-111	68	F68S	TTC->TCC	Gal^R
tfb3-88	68	F68V	TTC->GTC	MPA^S
tfb3-37	73	F73S	TTC->TCC	MPA^S/Gal^R
tfb3-7	88	V88E	GTG->GAG	Gal^R
tfb3-47	112	E112K	GAG->AAG	His^+
tfb3-1	112	E112V	GAG->GTG	His^+
tfb3-10	131	E131V	GAG->GTG	His^+
tfb3-155	204	D204N	GAT->AAT	His^+
tfb3-34	218	L218S	TTG->TCG	MPA^S

3.2.3 *tfb3* single mutants with transcriptional and/or conditional phenotypes are mainly clustered in the RING finger domain

Spots assay results for 31 single *tfb3* mutants are shown in **Figure 23** and summarized in **Table 6**. These results reveal that our *TFB3* screen has successfully identified two major classes of *tfb3* alleles that were predicted to alter TSS selection in both directions. One class of alleles conferred an MPA^S phenotype, meaning they were presumed to have initiation defects at *IMD2* and therefore are further predicted to shift TSS usage upstream. Another class of alleles conferred a His$^+$ phenotype, meaning they were able to constitutively express *imd2*Δ*::HIS3* and therefore are predicted to shift TSS usage downstream. No Spt$^-$ mutant was observed in the *TFB3* screen, which is different from *ssl2* mutants identified by our lab (Zhao et al., 2021).

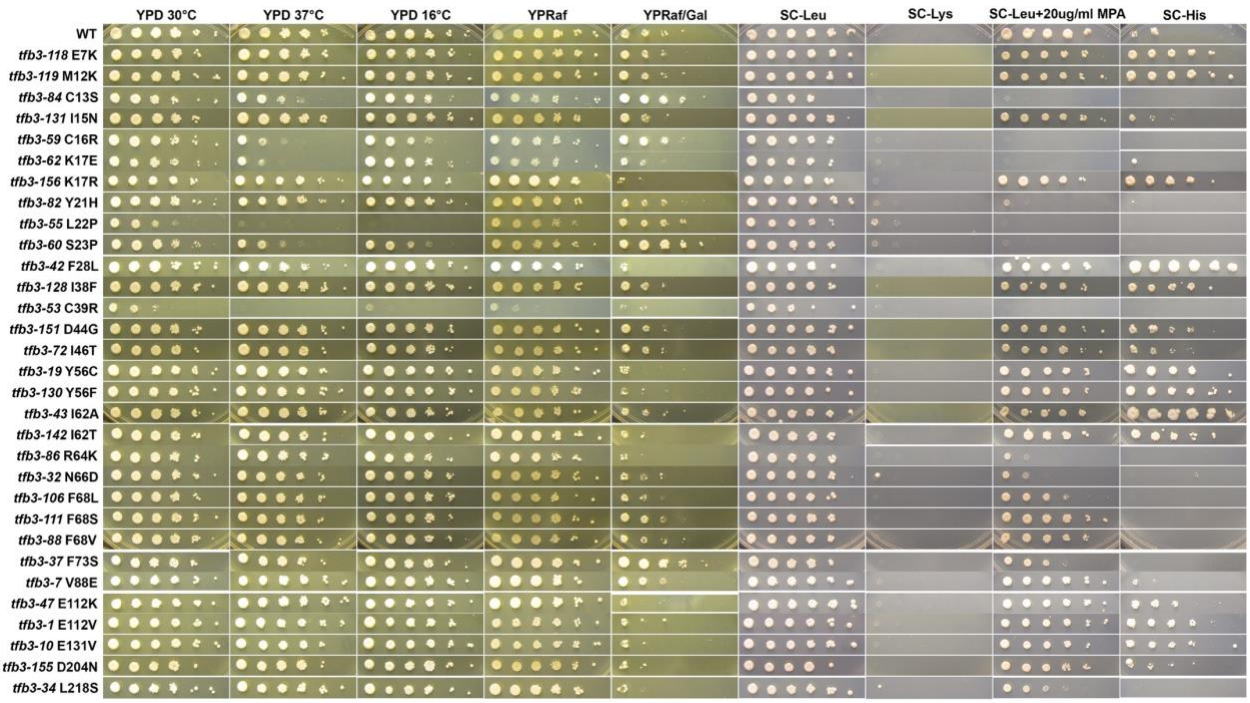


Figure 23 Spot assay of 31 *tfb3* single mutants

Growth phenotypes of *tfb3* single mutants on different media. Mutants are ordered based on substitution residues.

Table 6 Summary of spot assay results of 31 *tfb3* single mutants

Colored cells for phenotype columns indicate the strength of phenotypes, with increased color intensity representing increased level of phenotypes. Each phenotype has five ranked strengths from strong to weak: severe, strong, moderate, mild, and weak. The "structure information" column is based on Tfb3 structures and/or homolog amino acid in human MAT1 (Gervais et al., 2001).

***tfb3* allele #**	**AA substitution**	**Human MAT1 homolog**	**Structure information**	**Phenotype(s)**	**Growth Defect**	**His^+**	**MPA^S**	**Gal^R**	**Spt^-**	**Ts**	**Cs**
tfb3-118	E7K	/		His^+	0	4	0	0	0	0	0
tfb3-119	M12K	/		His^+	0	4	0	0	0	0	0
tfb3-84	C13S	C6	binding to Zn-I	MPA^S/Gal^R/ Ts	1	0	5	3	0	1	0
tfb3-131	I15N	R8		His^+	0	1	0	0	0	0	0
tfb3-59	C16R	C9	binding to Zn-I	MPA^S/Gal^R/ Ts/Cs	1	0	5	2	0	3	1
tfb3-62	K17E	K10		MPA^S/Ts	1	0	5	0	0	3	0
tfb3-156	K17R	K10		His^+	0	3	0	0	0	0	0
tfb3-82	Y21H	Y14	α1	MPA^S	0	0	4	0	0	0	0
tfb3-55	L22P	R15	α1	MPA^S/Gal^R/ Ts/Cs	3	0	5	1	0	4	5
tfb3-60	S23P	N16		MPA^S/Gal^R/ Ts/Cs	1	0	5	2	0	3	2
tfb3-42	F28L	L21	β1	His^+	0	4	0	0	0	0	0
tfb3-128	I38F	L30	β2; internal face of β sheet; hydrogen-bonded	His^+	0	3	0	0	0	0	0
tfb3-53	C39R	C31	β2; binding to Zn-I	MPA^S/Ts/Cs	4	0	5	0	0	5	4
tfb3-151	D44G	D36	α2; highly acidic AA located on helix	His^+	0	2	0	0	0	0	0
tfb3-72	I46T	L38	α2; highly conserved hydrophobic residues	His^+	0	2	0	0	0	0	0
tfb3-19	Y56C	/		His^+	0	3	0	0	0	0	0
tfb3-130	Y56F	/		His^+	0	3	0	0	0	0	0
tfb3-43	I62A	P52		His^+	0	4	0	0	0	0	0
tfb3-142	I62T	P52		His^+	0	3	0	0	0	0	0
tfb3-86	R64K	R54	highly positively charged surface	MPA^S	1	0	4	0	0	0	0
tfb3-32	N66D	S56		MPA^S	0	0	3	0	0	0	0
tfb3-106	F68L	F58	loop L2; highly conserved hydrophobic residues	MPA^S	0	0	2	0	0	0	0

Table 6 (continued).

tfb3-111	F68S	F58	loop L2; highly conserved hydrophobic residues	Gal^R	0	0	0	1	0	0	0
tfb3-88	F68V	F58	loop L2; highly conserved hydrophobic residues	MPA^S	0	0	1	0	0	0	0
tfb3-37	F73S	F63		MPA^S/ Gal^R	0	0	2	3	0	0	0
tfb3-7	V88E	V78		Gal^R	0	0	0	2	0	0	0
tfb3-47	E112K	E101		His^+	0	1	0	0	0	0	0
tfb3-1	E112V	E101		His^+	0	3	0	0	0	0	0
tfb3-10	E131V	E120		His^+	0	3	0	0	0	0	0
tfb3-155	D204N	P193		His^+	0	1	0	0	0	0	0
tfb3-34	L218S	Q207		MPA^S	0	0	2	0	0	0	0

Mapping substituted residues of 31 singles onto the Tfb3 primary structure shows that most substitutions were located in the Tfb3 RING finger domain (**Figure 24**). Though two substitutions within the α-helical domain showed His^+ and MPA^S phenotypes, their phenotype strengths are very weak, relative to the RING finger domain substitutions. These observations suggest that Tfb3 functions in TSS selection process mainly via its RING finger and ARCH anchor domains. One interesting position is K17, where two substitutions at this residue showed opposite and strong transcription-related phenotypes, namely K17E caused a severe MPA^S phenotype and K17R caused a moderate His^+ phenotype. These data suggest a potentially important role for this residue.

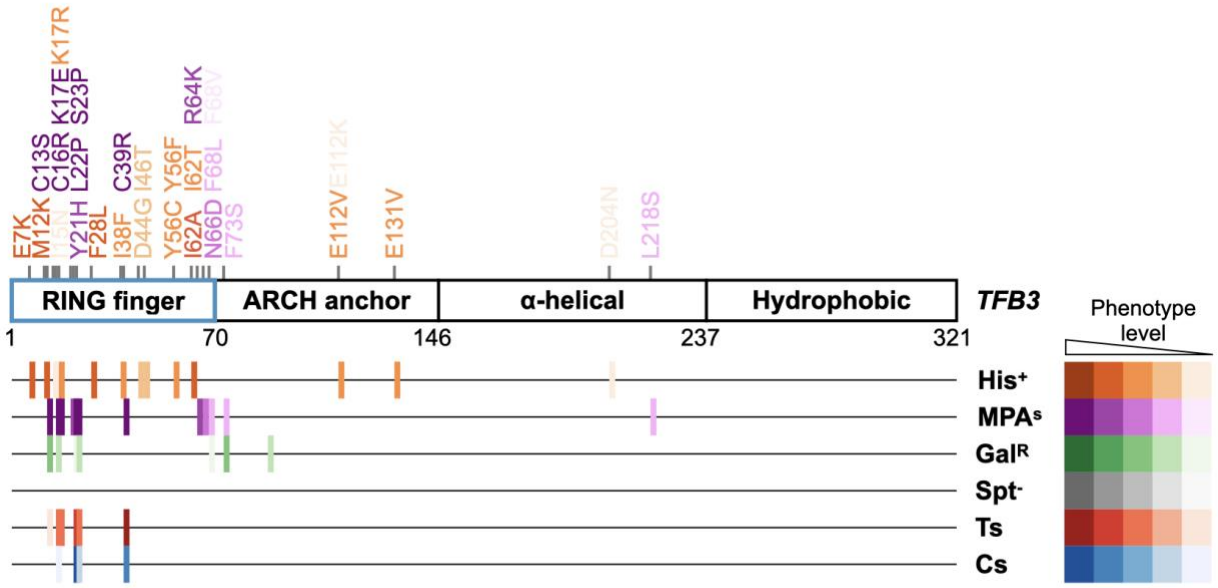


Figure 24 Mapping identified *tfb3* single mutants to Tfb3 primary structure

Colored lines in the lower panel indicate phenotypes of each mutant with colors representing the strength of phenotypes. Single mutants conferring MPA^S or His^+ phenotype are labelled in the upper panel and colored based on their His^+ or MPA^S phenotype.

To investigate how these substitutions and corresponding phenotypes are related to their residue locations within the protein structure and/or involved in potential networks with other proteins, we mapped substitutions to the Tfb3 protein structure (**Figure 25**). The 13 substitutions at 12 residues conferring the MPA^S phenotype are mainly located in three clusters. The first cluster of substitutions (R64K, N66D) are located at the RING finger interface with Pol II Rpb4-Rpb7 stalk subcomplex (**Figure 25**, substitutions in purple in top middle panel). The second cluster of substitutions (K17E, Y21H, L22P, and S23P) are located at the RING finger interface with TFIIE subunit Tfa1 E-linker helices (**Figure 25**, substitutions in purple in bottom right panel). The third cluster of substitutions (C13S, C16R, C39R, F68L, and F68V) are residues that bind to zinc ion I (Zn-I) or are located surrounding the zinc binding residues (**Figure 25**, substitutions in purple in

bottom left panel). In addition, the mild MPA^S allele F73S is located at the surface of the ARCH domain (**Figure 25**, top right panel). The region in which the mild MPA^S mutant L218S is located has not been resolved in any structure yet. The 16 substitutions at 13 residues conferring the His^+ phenotype are mainly located at internal residues within the RING finger domain (**Figure 25,** substitutions in orange in bottom left panel), most of which may increase stability or alter dynamics of the RING finger domain or help zinc binding. For example, F28 and I38 are located in internal β1 and β2 sheets and predicted to contribute to αββαβ fold of the RING finger domain (Gervais et al., 2001). M12 and I15 are two hydrophobic residues adjacent to Zn-I binding residues C13 and C16, respectively, and might help stable zinc binding and therefore contribute to stabilizing the RING finger core.

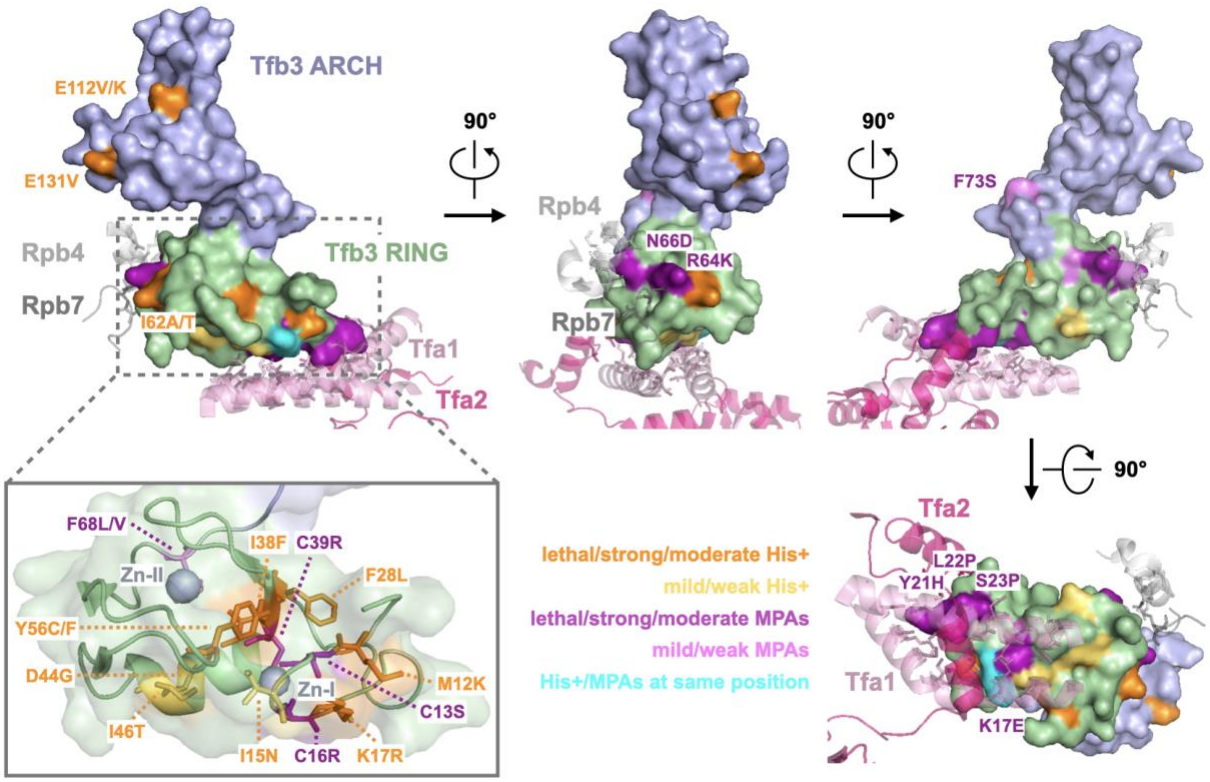


Figure 25 Mapping identified *tfb3* subsitutions to Tfb3 protein structure

Identified substituted residues were mapped to Tfb3 protein structure (PDB: 7O4L) (Schilbach et al., 2021). The RING finger (in pale green) and ARCH anchor (in light blue) domains of Tfb3 are depicted as surface. Rpb4-Rpb7 stalk (in gray90 and gray50, respectively) and TFIIE (subunit Tfa1 in pink and Tfa2 in hotpink) are depicted as cartoon, with residues within interface with Tfb3 shown as sticks as well. Amino acid substitutions causing His^+ phenotype are colored in orange (strong or moderate His^+) or yellow orange (mild or weak) and causing MPA^S phenotype are colored in purple (severe, strong or moderate MPA^S) or violet (mild, weak). Positions that can cause either His^+ or MPA^S phenotype depending on the particular substitutions are colored in cyan. A close-up and internal view shows RING finger domain structure as cartoon (in pale green) and two zinc atoms as spheres (in light blue). Residues at which substitutions showed His^+ or MPA^S phenotypes are depicted as sticks as well.

3.2.4 *tfb3* single mutants show intermediate TSS usage effects between "efficiency" and "processivity" mutants

To quantitatively determine TSS usage effects of identified *tfb3* mutants conferring MPA^S or His^+ phenotype, TSS usages at *ADH1* were examined by Primer Extension (PE). *ADH1* contains two major TSSs (-37 nt and -27 nt upstream of start codon) and some minor TSSs (**Figure 26A**), making it be a good model gene for studying effects on TSS distributions. Additionally, investigation of TSS usage at *ADH1* can be used as a proxy for potential global defects as it is a gene unrelated to *IMD2*, which was used for *TFB3* screen. Primer extension was performed on representative *tfb3* single mutants with a range of MPA^S or His^+ phenotype strengths and locations in different domains of Tfb3 (**Figure 26**). The quantification strategy for detected *ADH1* TSS usage is as described in (Jin and Kaplan, 2014) (**Figure 26A**). Consistent with previously studied Pol II and other GTF mutants, all tested *tfb3* MPA^S alleles shifted TSS usage upstream at *ADH1* (**Figure 26B)**, whereas *tfb3* His^+ alleles shifted TSS usage downstream (**Figure 26C)**. Interestingly and provocatively, the pattern of TSS usage changes of *tfb3* MPA^S mutants was distinct from Pol II efficiency mutants (**Figure 26D)** and *ssl2* processivity mutants (**Figure 26E)**. Instead, *tfb3* alleles showed an intermediate pattern between efficiency and processivity mutants. Specifically, Pol II efficiency upstream shifters (e.g., E1103G) shift TSS distribution upstream by increasing the usage of upstream minor TSSs, resulting in increased TSS usage in the bin-2 of our quantification scheme (**Figure 26D)**. In contrast, *ssl2* processivity upstream shifters (e.g., N230D) shift TSS distribution upstream by decreasing the usage of downstream TSSs rarely affecting upstream minor TSSs, resulting in primarily increasing usage in the bin-3 TSS (**Figure 26E)**. However, *tfb3* upstream shifters showed increased usages in both bin-2 and bin-3 in quantification (**Figure 26B)**, which is intermediate between tested Pol II and *ssl2* alleles.

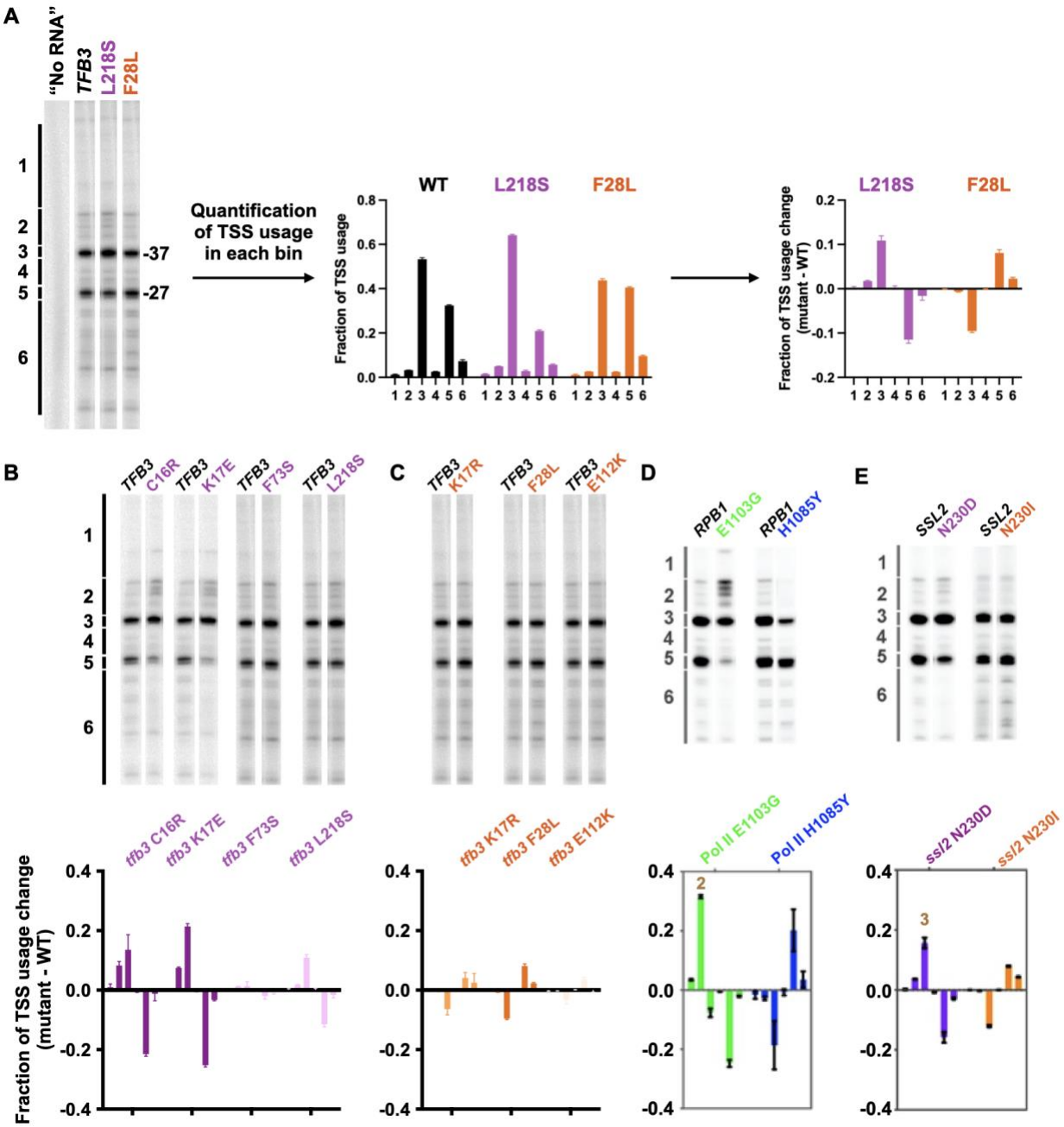


Figure 26 Primer extension and quantification of *tfb3* single mutant effects on *ADH1* TSS usage

(A) Quantification strategy for detected *ADH1* TSS usage. *ADH1* TSS signals were divided into six bins, quantified, and normalized to total signal from all bins, with background removed based on "no RNA" reaction signals. The change of TSS usage fraction for mutants compared to WT was calculated by subtracting TSS usage of each bin in WT from that in mutants. **(B-E)** Comparison of *ADH1* TSS usages in WT, analyzed *tfb3* mutants, and representative Pol II and *ssl2* mutants. Upper panels are TSSs detected by primer extension. Bottom panels show quantification of TSS usage change of mutants from WT. *tfb3* MPA^S **(B)** and His$^+$ **(C)** mutants are, respectively, colored in purple and

orange, with increased color intensity representing increased phenotype strength. **(D-E)** Pol II and *ssl2* mutants primer extension data are from (Zhao et al., 2021). **(D)** Pol II upstream and downstream shifters are colored in green and blue, respectively. **(E)** *ssl2* upstream and downstream shifters are colored in purple and orange, respectively.

To further investigate Tfb3 function in transcription initiation, we examined *tfb3* genetic interactions with mutants within other initiation factors such as Ssl2 and the PIC cofactor Sub1. Each of these genetic interaction studies will be discussed below.

3.2.5 *tfb3* alleles show additive interactions with *ssl2* alleles

Ssl2, the ATP-dependent translocase/helicase of TFIIH, promotes DNA melting and modulates promoter scanning by translocating DNA into the Pol II active site and by doing so, determines the scanning window within which Pol II activity will determine initiation efficiency (Grunberg et al., 2012; Zhao et al., 2021). The "Latch region" (residues 120-314) of Tfb3, a proposed structural motif, has been observed to crosslink to the Lock-N region of Ssl2, as well as other TFIIH subunits such as Rad3 and Ssl1, and therefore anchors the TFIIH kinase to core modules (Luo et al., 2015; Robinson et al., 2016; van Eeuwen et al., 2021). This raises the possibility that Tfb3 might function through TFIIH during scanning. However, recent structural data indicate that these interactions are rearranged within the PIC during initiation. That being said, how Tfb3 activity cooperates or supports Ssl2 function is unknown, therefore we investigated this by examining genetic interaction between *tfb3* alleles and *ssl2* alleles. We generated *tfb3/ssl2* double mutants and examined their effects on growth (**Figure 27**) and transcriptional phenotypes (discussed later). Remarkably, extensive synthetic lethality or strongly synthetic sickness was observed between *tfb3* MPA^S alleles with *ssl2* MPA^S alleles (**Figure 27**, in black box), indicating

additive interactions between *tfb3* and *ssl2* alleles. Interestingly, *tfb3* mild MPA^S allele F73S showed synthetic lethality with two *ssl2* strong MPA^S alleles (V473D and D522V) and one mild allele (Y750*); however, only showed synthetic sickness but not lethality with *ssl2* strong MPA^S allele N230D. This might suggest some special interactions between these particular positions of Tfb3 and Ssl2, or functional interaction between domains these residues are located, namely the Tfb3 ARCH anchor domain (F73S) and two helicase domains of Ssl2 (V473D and D522V in HD1, Y750* in HD2). Moreover, *ssl2* His$^+$ alleles N230I and R636C showed some level of suppression of growth defects of *tfb3* MPA^S alleles L22P and C39R (**Figure 27**, in yellow boxes), consistent with the additive interaction hypothesis.

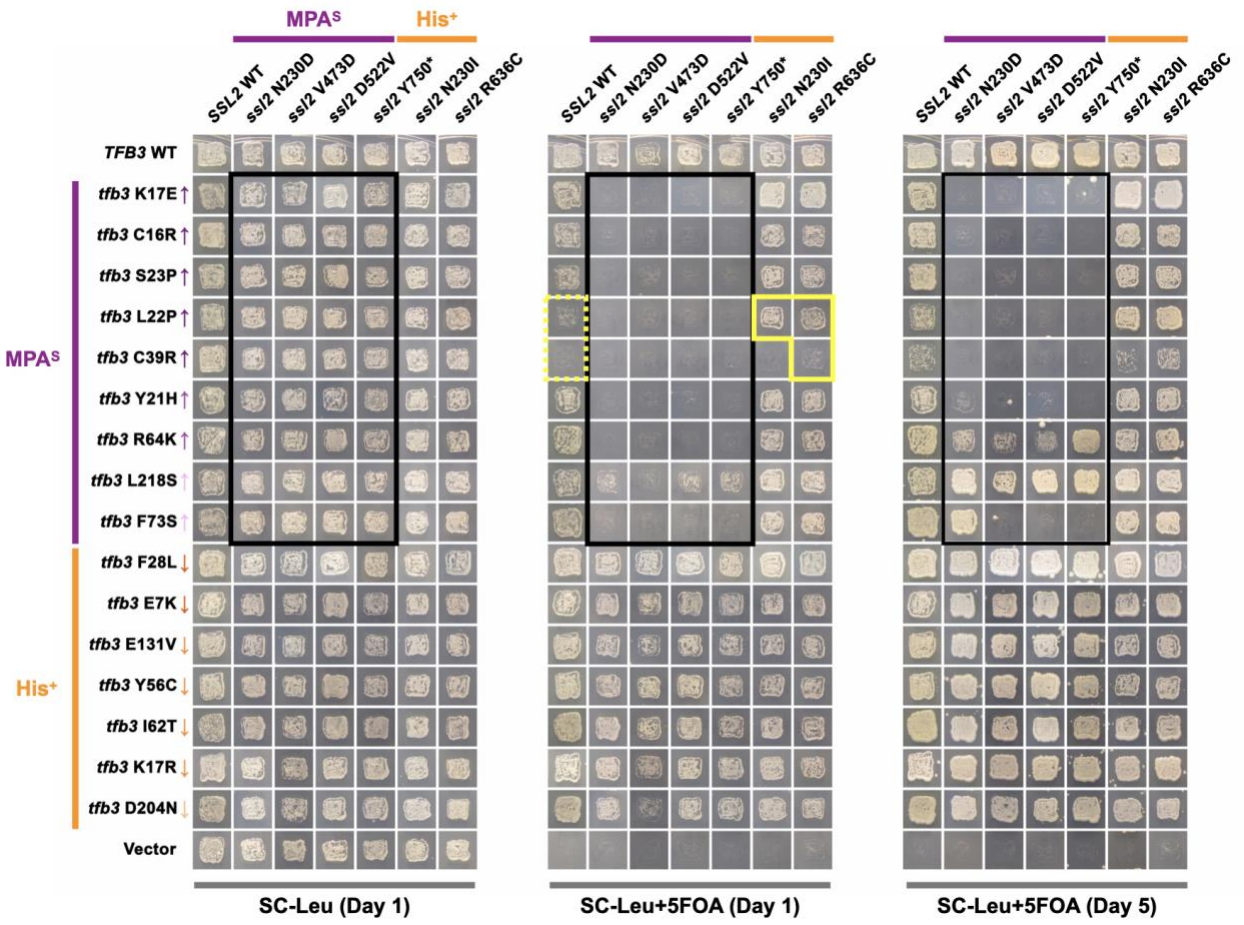


Figure 27 Growth phenotypes of *tfb3/ssl2* double mutants

Viability of *tfb3/ssl2* double mutants were assayed by plasmid shuffling. The left panel shows growth on SC-Leu plate which is the control state where WT *TFB3* is present. Middle and right panels show growth on SC-Leu+5FOA plate on day 1 and day 5, respectively, indicating growth phenotypes of double mutants. Up and down arrows indicate PE detected or predicted single mutant effects on TSS distribution, with increased color intensity representing increased strength of phenotypes. Patches in black boxes are double mutants that combine *tfb3* MPA^S mutants and *ssl2* MPA^S mutants. Patches in yellow solid boxes are double mutants that combine *tfb3* MPA^S mutants with growth defect and *ssl2* His+ mutants, with patches of corresponding *tfb3* single mutants in the yellow dashed box.

To examine *tfb3/ssl2* double mutant effects on TSS usage and therefore assess how *tfb3* alleles genetically interact with *ssl2* alleles, we selected K17E (MPA^S, TSSs upstream shifter) and F28L (His^+, TSSs downstream shifter) as representative of the two *tfb3* allele classes and performed spot assays on the viable *tfb3/ssl2* double mutants (**Figure 29A, Figure 30A, Figure 31A**). Additionally, we did preliminary phenotype investigation for other double mutants by replicating patch assay plates onto phenotyping plates (**Figure 28, Figure 29B, C, Figure 30B, Figure 31B**). We observed primarily additive/suppressive interactions between *tfb3* and *ssl2* alleles, suggesting Tfb3 is potentially involved in scanning processivity based on our understanding of Ssl2 function (Zhao et al., 2021). Each double mutant class based on phenotypes of singles will be discussed below.

(1) *tfb3* MPA^S (up) + *ssl2* MPA^S (up) = synthetic lethality or sickness. As shown in viability assay (**Figure 27**), *tfb3* MPA^S mutants (PE at *ADH1* detected or predicted upstream shifting alleles) show synthetic lethality or strong growth defects when combined with *ssl2* MPA^S mutants (TSS-seq measured and/or PE detected upstream shifting alleles), suggesting additive interactions between these alleles. Furthermore, we replicated patches onto phenotyping plates to preliminarily investigate for MPA^S and His^+ phenotypes of viable double mutants. As shown in **Figure 28**, additive interactions were observed. Especially, the *tfb3* L218S/*ssl2* Y750* double mutant, which combines two weak MPA^S mutants, showed a stronger MPA^S phenotype (**Figure 28**, in purple solid box) than corresponding *tfb3* L218S (**Figure 23**) and *ssl2* Y750* (**Figure 28**, in purple dashed box) single mutants.

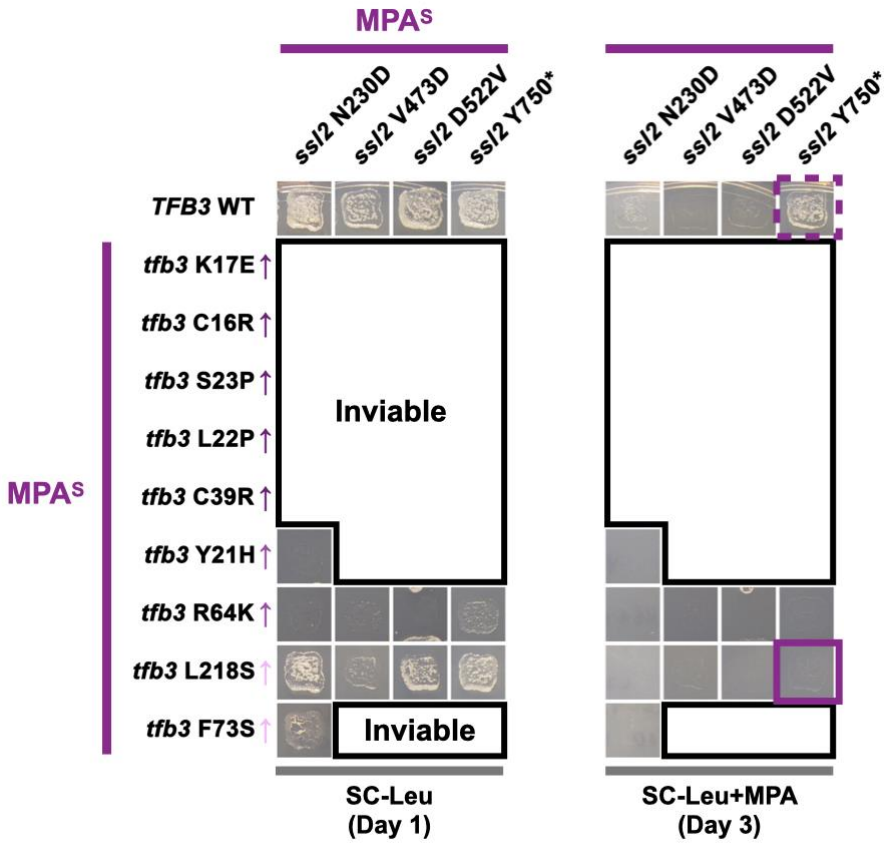


Figure 28 Transcription-related phenotypes of *tfb3* MPA^S/*ssl2* MPA^S double mutants

Patch assay plates used in **Figure 27** were replicated onto phenotyping plates (SC-Leu+20 μg/ml MPA) and control plate (SC-Leu). Patch in the purple solid box is the double mutant combining *tfb3* weak MPA^S L218S allele and *ssl2* weak MPA^S Y750* allele, with the patch of corresponding *ssl2* single mutant in the purple dashed box.

(2) ***tfb3* MPA^S (up) + *ssl2* His$^+$ (down) = additivity**/suppression. *tfb3* MPA^S alleles showed additive interactions with *ssl2* His$^+$ alleles, meaning both MPA^S phenotype of *tfb3* alleles and His$^+$ phenotype of *ssl2* alleles were mutually suppressed to some extent in *tfb3/ssl2* double mutants (**Figure 29)**. When combining the *tfb3* severe MPA^S allele K17E (**Figure 23)** with the *ssl2* strong His$^+$ allele N230I, weaker MPA^S and weaker His$^+$ phenotypes were observed in the double mutant (**Figure 29A**), indicating an additive interaction (suppression through addition of opposite behaving mutants). In addition, the strong Spt$^-$ phenotype of *ssl2* N230I was partially

suppressed and showed a weak Spt$^-$ phenotype in the double mutant (**Figure 29A**, in black boxes), consistent with an additive interaction hypothesis. The *tfb3* K17E allele apparently showed epistasis to the *ssl2* moderate His$^+$ allele R636C, meaning the His$^+$ phenotype of *ssl2* R636C was completely suppressed by *tfb3* K17E, and the double mutant showed severe MPA^S phenotype that was close to the *tfb3* K17E single mutant. However, because phenotypes may have a threshold effect, epistasis can be apparent if additive interactions are not enough to get over threshold. This could potentially be addressed by a more sensitive MPA assay using lower MPA amounts. If additive, a less degree of MPA^S phenotype of the *tfb3* K17E/*ssl2* R636C double mutant than the *tfb3* K17E single mutant is expected. In contrast, if *tfb3* K17E is epistatic to *ssl2* R636C, the same MPA^S level as the *tfb3* K17E allele is expected even at lower MPA amounts. While whether the interaction between *tfb3* K17E and *ssl2* R636C is additive or epistatic is not certain yet, the broad additivity/suppression of transcriptional phenotypes were observed in patch phenotyping assay for additional double mutants combining *ssl2* His$^+$ alleles (N230I and R636C) with *tfb3* mutants with different strengths of MPA^S phenotype (**Figure 29B, C**). Specifically, strong but to a lesser degree of MPA^S phenotypes than *tfb3* single alleles were observed in combination of *tfb3* severe MPA^S alleles (C16R, S23P, L22P, C39R) with *ssl2* strong His$^+$ N230I (**Figure 29B**, in purple solid box). Additionally, all double mutants showed His$^+$ phenotypes (**Figure 29C**, in orange solid box) with His$^+$ strength weaker than *ssl2* single mutants (**Figure 29C**, in orange dashed box). Taken together, we concluded that *tfb3* MPA^S alleles showed additive interactions with *ssl2* His$^+$ alleles.

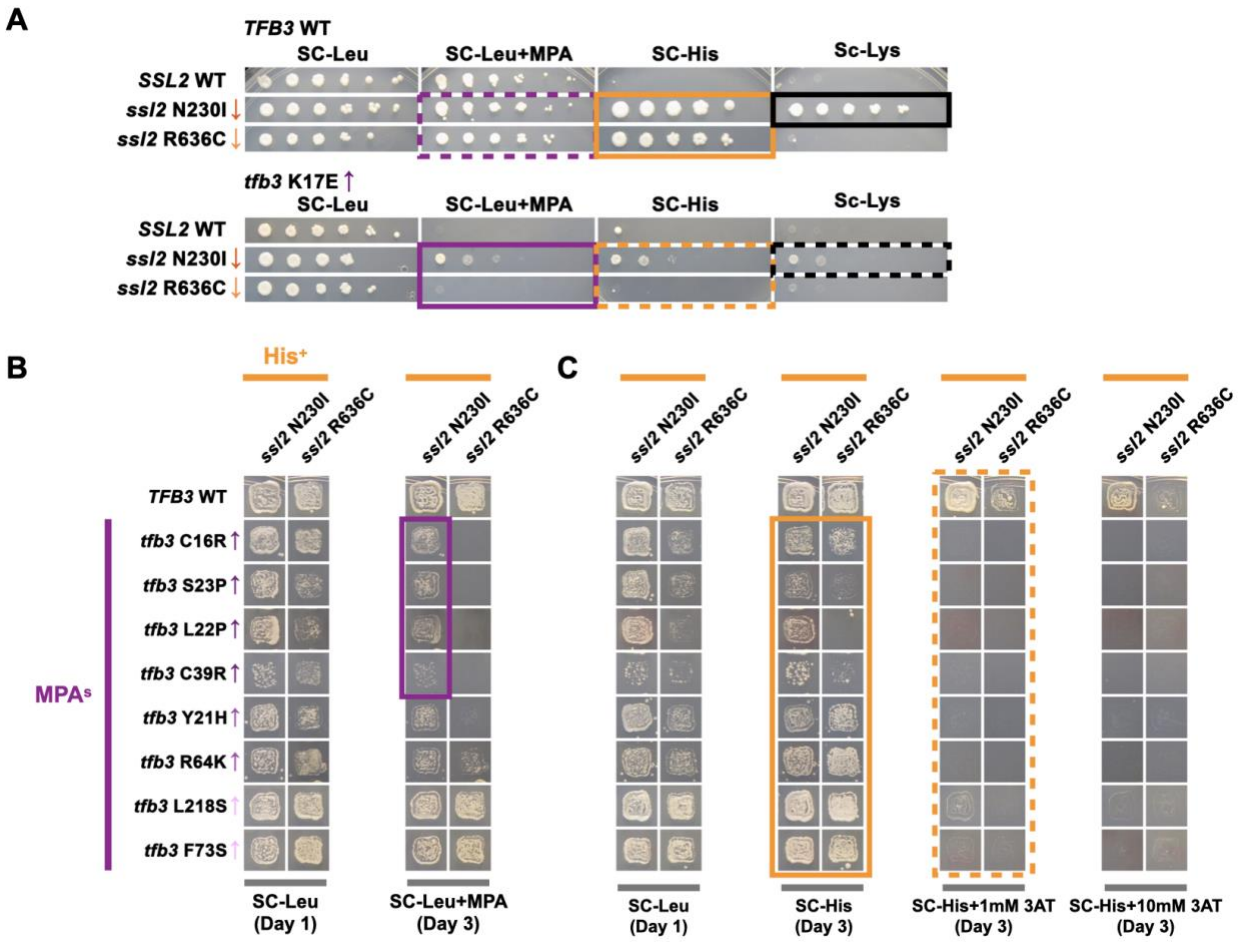


Figure 29 Transcription-related phenotypes of *tfb3* MPA^S/*ssl2* His$^+$ double mutants

(A) Spot assay results for double mutants combining *tfb3* K17E (MPA^S allele) and *ssl2* His$^+$ alleles. Double mutants showed strong MPA^S phenotypes (in the purple solid box), which was not observed for *ssl2* single mutants (in the purple dashed box). Meanwhile, double mutants showed no or much weaker His$^+$ phenotypes (in the orange dashed box) than *ssl2* single mutants (in the orange solid box). In addition, the strong Spt$^-$ phenotype of *ssl2* N230I (in the black solid box) was partially suppressed (in the black dashed box) by *tfb3* K17E. **(B and C)** Patch phenotyping assay of *tfb3* MPA^S/*ssl2* His$^+$ double mutants. Patch assay plates used in **Figure 27** were replicated onto phenotyping plates (SC-Leu+20 μg/ml MPA, SC-His, SC-His+1 mM 3AT, SC-His+10 mM 3AT) and control plate (SC-Leu). Arrows beside *tfb3* single mutants indicate PE detected or predicted TSS shifting direction (upstream shifting for MPA^S), with increased color intensity representing increased MPA^S phenotype strength.

(3) *tfb3* His$^+$ (down) + *ssl2* MPA^S (up) = additivity/suppression. *tfb3* His$^+$ alleles showed putatively additive interactions with *ssl2* MPA^S alleles, meaning the MPA^S phenotype of *ssl2* alleles were suppressed to some extent in *tfb3/ssl2* double mutants (**Figure 30**). When combining the *tfb3* strong His$^+$ allele F28L with *ssl2* strong or moderate MPA^S alleles, a lesser degree of MPA^S phenotypes than *ssl2* single alleles were observed in double mutants (**Figure 30A**), suggesting additive interactions. While we were not able to construct double mutants in the genetic background of the *imd2*Δ*::HIS3* reporter, this additive interaction was supported by additional double mutants combining *ssl2* MPA^S alleles with *tfb3* alleles with different strengths of His$^+$ phenotype (**Figure 30B**). Specifically, double mutants combining *tfb3* strong or moderate His$^+$ alleles with *ssl2* MPA^S alleles showed weaker MPA^S phenotypes (**Figure 30B**, in purple dashed box) compared to corresponding *ssl2* single mutants (**Figure 30B**, in purple dashed box). This additivity hypothesis could be further confirmed by investigating whether double mutants combining strong *tfb3* His$^+$ alleles with weak *ssl2* MPA^S alleles, such as *tfb3* F28L/*ssl2* Y750* and *tfb3* E7K/*ssl2* Y750*, also show His$^+$ phenotype. Taken together, we speculate *tfb3* His$^+$ alleles show additive interactions with *ssl2* MPA^S alleles.

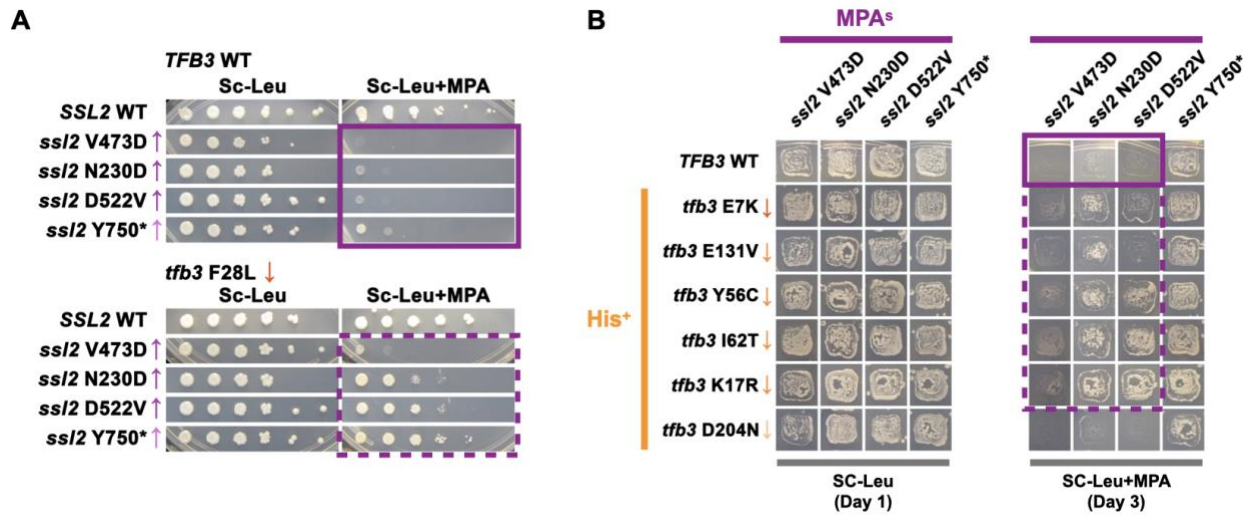


Figure 30 Transcriptional phenotypes of *tfb3* His$^+$/*ssl2* MPA^S double mutants

(A) Spot assay results for double mutants combining the *tfb3* His$^+$ allele F28L and *ssl2* MPA^S alleles. Double mutants showed weaker MPA^S phenotypes (in purple dashed box) than *ssl2* single mutants (in purple solid box). Arrows beside *ssl2* single mutants indicate PE detected or predicted TSS shifting direction (upstream shifting for MPA^S), with increased color intensity representing increased MPA^S phenotype strength. **(B)** Patch phenotyping assay of *tfb3* His$^+$/*ssl2* MPA^S double mutants. Patch assay plates used in **Figure 27** were replicated onto phenotyping plate (SC-Leu+20 µg/ml MPA) and control plate (SC-Leu). Arrows beside *tfb3* single mutants indicate PE detected or predicted TSS shifting direction (downstream shifting for His$^+$), with increased color intensity representing increased His$^+$ phenotype strength.

(4) *tfb3* His$^+$ (down) + *ssl2* His$^+$ (down) = additivity or epistasis. *tfb3* His$^+$ alleles showed additive or epistatic interactions with *ssl2* His$^+$ alleles. When combining the *tfb3* strong His$^+$ F28L allele with *ssl2* strong or moderate His$^+$ alleles N230I or R636C, stronger or at least the same level of His$^+$ phenotypes as *ssl2* single mutants were observed in double mutants (**Figure 31A**). Because we don't have data for SC-His medium with 3-AT (3-Amino-1,2,4-triazole, a competitive inhibitor of *HIS3* gene product) for *tfb3* F28L single mutant and double mutants, we are not able to compare the strength of His$^+$ phenotypes of *tfb3* F28L single and double mutants. Therefore, we are not able to distinguish whether *tfb3* His$^+$ alleles are additive or epistatic to *ssl2* His$^+$ alleles at the present.

This could be addressed by examining His^+ strength of all single and double mutants using higher 3-AT. More importantly, performing primer extension will allow us to quantitatively examine mutant effects on TSS distribution. Though further investigation is needed, combination of the *ssl2* R636C with other *tfb3* His^+ alleles showed stronger His^+ phenotypes than *ssl2* R636C single (**Figure 31B**, in orange solid boxes), consistent with the hypothesis that *tfb3* His^+ alleles were putatively additive or epistatic to *ssl2* His^+ allele.

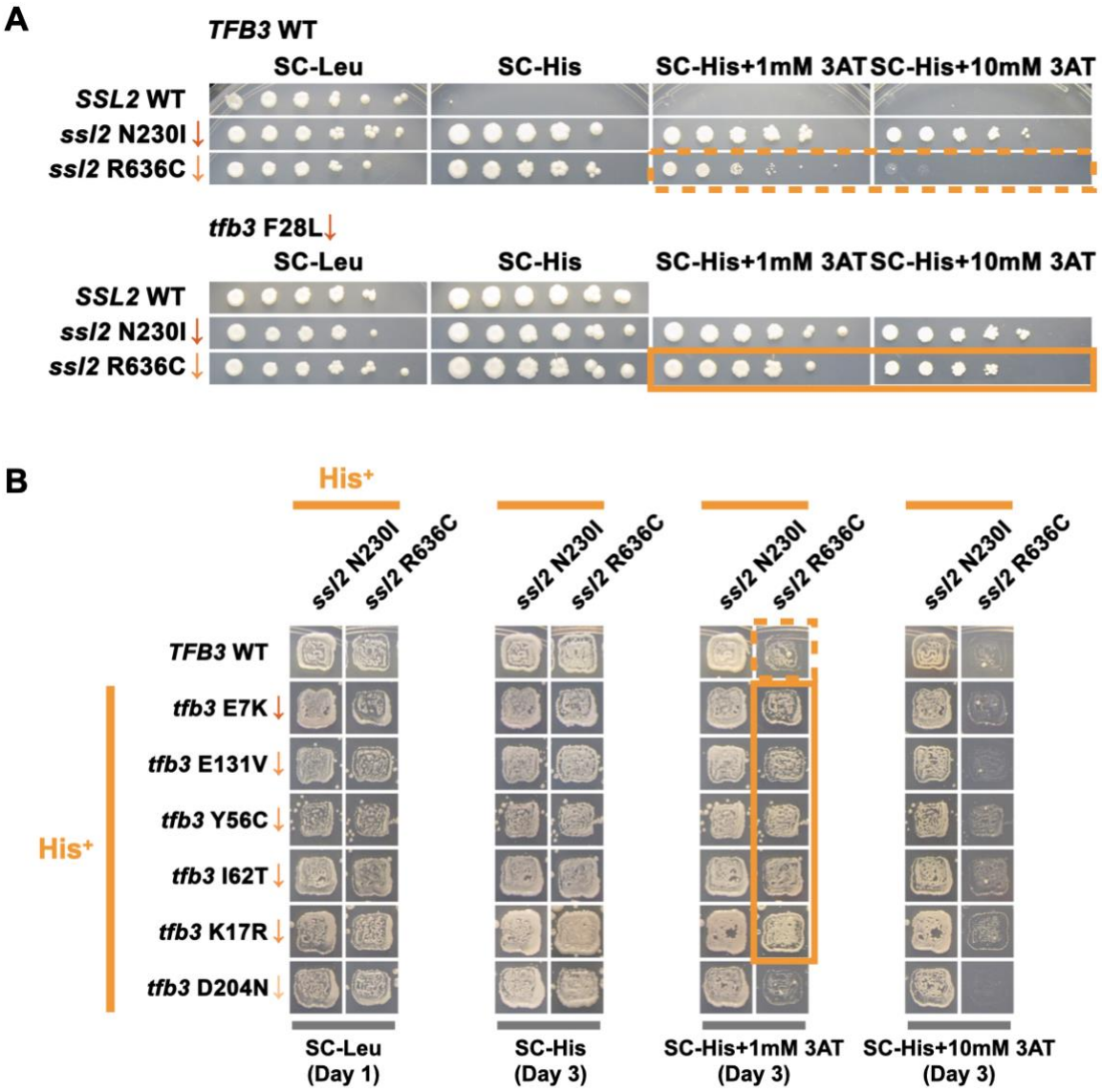


Figure 31 Transcriptional phenotypes of *tfb3* His$^+$/*ssl2* His$^+$ double mutants

(A) Spot assay results for double mutants combining the *tfb3* F28L (His$^+$ allele) and *ssl2* His$^+$ alleles. The *tfb3* F28L/*ssl2* R636C double mutant showed stronger His$^+$ (in orange solid box) than *ssl2* R636C single mutant (in orange dashed box). Arrows beside *ssl2* single mutants indicate TSS-seq detected TSS shifting direction (downstream shifting), with increased color intensity representing increased His$^+$ phenotype strength. **(B)** Patch phenotyping assay of *tfb3* His$^+$/*ssl2* His$^+$ double mutants. Patch assay plates used in **Figure 27** were replicated onto phenotyping plate (SC-His, SC-His+1 mM 3AT, SC-His+10 mM 3AT) and control plate (SC-Leu). Arrows beside *tfb3* single mutants indicate PE detected or reporter predicted TSS shifting direction (downstream shifting for His$^+$), with increased color intensity representing increased His$^+$ phenotype strength.

3.2.6 *tfb3* alleles show additive interactions with *sub1*Δ

Sub1 is a PIC component (Sikorski et al., 2011) and physically interacts with the junction between single- and double-stranded DNA (Sikorski et al., 2011) and multiple PIC subunits (Garavis et al., 2017; Knaus et al., 1996). Deletion of *SUB1* (*sub1*Δ) itself significantly shifts TSS distribution downstream (Braberg et al., 2013; Koyama et al., 2008), suggesting Sub1 contributes to TSS selection. Furthermore, its distinct genetic interactions with Pol II or *sua7-1* alleles compared with *ssl2* alleles suggested that Sub1 might be involved in the processivity network (including Ssl2) instead of the efficiency network (including Pol II, TFIIB, and TFIIF) controlling initiation (Braberg et al., 2013; Jin and Kaplan, 2014; Zhao et al., 2021). If Tfb3 is involved in the processivity network as suggested by previous results, we predicted *tfb3* alleles to show additive genetic interaction with *sub1*Δ. To investigate genetic interaction between *tfb3* alleles and *sub1*Δ, we selected six representative *tfb3* alleles, including four MPA^S alleles (C16R, K17E, C39R, and F73S) and two His^+ alleles (K17R and F28L), constructed double mutants, and examined their transcriptional phenotypes (**Figure 32**). First, no strong growth defect (synthetic lethality) was observed between *tfb3* alleles and *sub1*Δ, even when combining alleles that both shift TSS distribution downstream. This observation was different from the broad synthetic lethality observed for combinations of two downstream shifters that are from different networks, such as Pol II LOF/*sub1*Δ (Braberg et al., 2013), *sua7 allele/sub1*Δ (Knaus et al., 1996), and *sua7-1/ssl2* GOF (Zhao et al., 2021). This suggested Tfb3 and Sub1 might function at the same step or affect the same process so that there is no further effect when combined. Second, when combined with *sub1*Δ, MPA^S phenotypes of *tfb3* alleles were suppressed to a certain extent depending on the original MPA^S phenotype strength (**Figure 32A**, in purple boxes). For example, the double mutant combining the *tfb3* mild MPA^S F73S allele and *sub1*Δ showed no MPA^S and mild His^+ phenotypes,

supporting additivity in their interaction. Third, double mutants combining alleles that both shift TSSs downstream showed ultra-strong His^+ phenotypes. However, limited by the threshold of spot assay on the His^+ phenotype, we cannot distinguish whether *sub1*Δ showed additive interactions or epistasis with *tfb3* His^+ alleles. Therefore, we performed primer extension to quantitatively examine mutant effects on TSS distribution, and we observed primarily additive interactions (**Figure 32B-D**). When strong *tfb3* upstream shifters, such as C16R and K17E, combined with *sub1*Δ (strong downstream shifter), double mutants showed close to WT TSS distribution (**Figure 32C**). It should be noted that the increased usage of bin-2, containing upstream minor TSSs, was still observed in these double mutants to some degree (**Figure 32B**, in purple box). When *tfb3* downstream shifters were combined with *sub1*Δ, usage in upstream bin-3 (major -37 TSS) was further decreased, whereas usage in downstream bin-6 was further increased, compared to either *tfb3* or *sub1*Δ single alleles (**Figure 32D**). Taken together, two types of *tfb3* alleles, including upstream shifters (MPA^S) and downstream shifters (His^+), showed additive interactions with *sub1*Δ, suggesting Tfb3 functions in TSS selection within the same processivity network as Sub1.

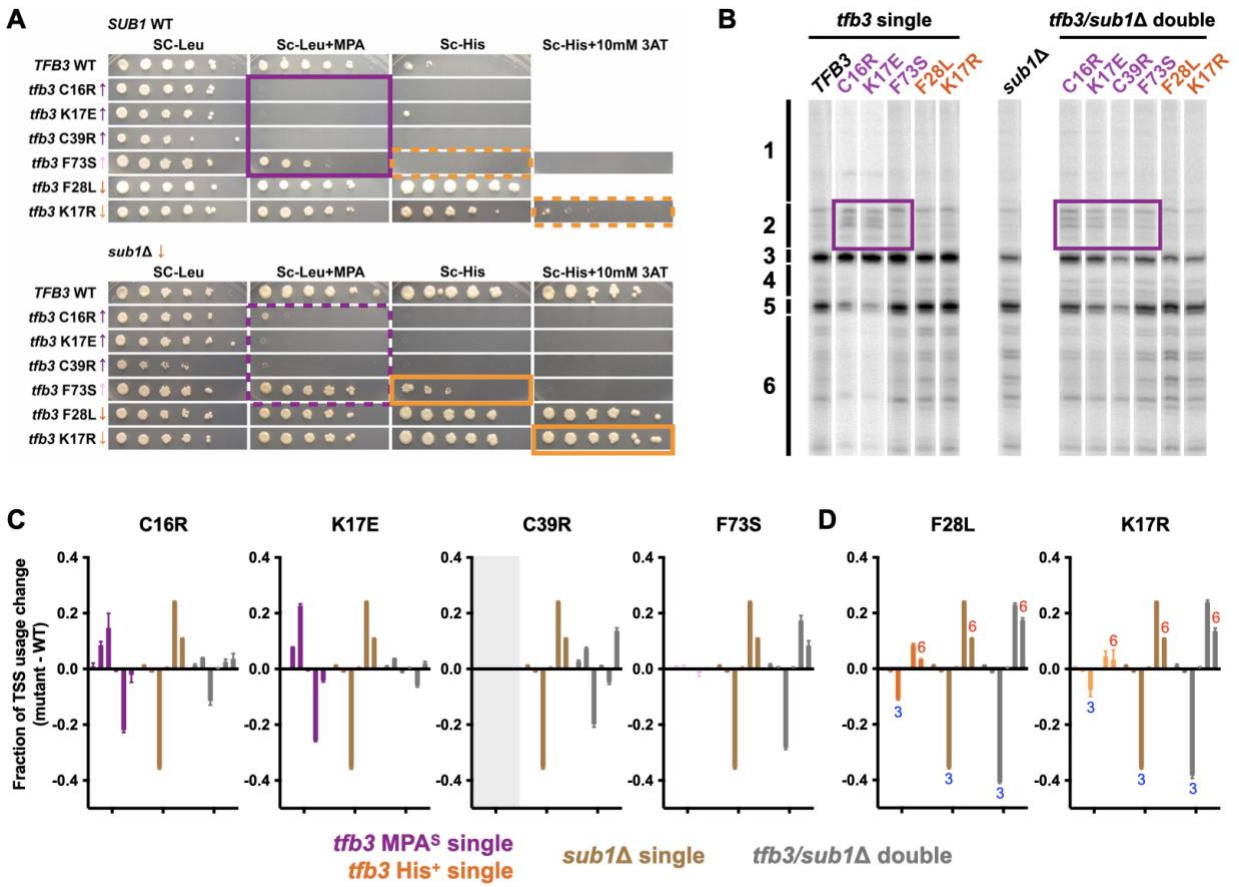


Figure 32 Additive genetic interaction between *tfb3* alleles and *sub1*Δ

(A) Spot assay results for double mutants combining *tfb3* alleles and *sub1*Δ. Spots in purple boxes indicate weaker MPA^S phenotype of doubles than *tfb3* strong MPA^S single alleles when combined with *sub1*Δ. Spots in orange boxes in the SC-His column indicate no MPA^S and weak His$^+$ phenotypes of doubles when a *tfb3* mild MPA^S single allele was combined with *sub1*Δ. Spots in orange boxes in SC-His+10 mM 3AT column indicate stronger His+ phenotype of doubles than *tfb3* moderate His$^+$ single allele when combined with *sub1*Δ. **(B-D)** Comparison of *ADH1* TSS usage in WT, *tfb3* single mutants, *sub1*Δ single mutant, and *tfb3*/*sub1*Δ double mutants. The same definition of bins of *ADH1* TSSs, quantification, normalization, and visualization methods as **Figure 26** have been used. **(B)** TSSs at *ADH1* detected by primer extension. **(C-D)** Quantification of TSS usage changes of single and double mutants from WT. Signals of *tfb3* MPA^S and His$^+$ alleles are in purple and orange, respectively, with increased color intensity representing increased phenotype strength. Signals of *sub1*Δ mutant are in brown. Signals of *tfb3/sub1*Δ double mutants are in gray. Primer extension result of *tfb3* C39R single mutant is missing so shadowed in grey.

3.3 Discussion

Tfb3 bridges TFIIH, Pol II, and TFIIE in the PIC and links the TFIIH kinase complex (TFIIK) to core TFIIH during RNA Pol II transcription initiation. A recent human PIC structural study has observed that the contact between TFIIH and the rest of the PIC through human MAT1 (Tfb3 homolog) changed during initiation and suggested this change can have consequences on TFIIH's activity (Aibara et al., 2021). This idea led us to propose that Tfb3 in yeast, which is highly conserved with its human homolog, MAT1, has similar functions during transcription initiation, but the contact change happens later in the scanning process that yeast uses to identify TSSs downstream from the core promoter. As a result, yeast TFIIH can support scanning downstream for TSSs (40-120 bp downstream of a TATA-box for TATA-containing promoters) in contrast to the human system where TSSs used are within the region of initial DNA melting (31-32 bp downstream of a TATA-box for TATA-containing promoters) and do not appear to require scanning. In support of this hypothesis, deletion of TFIIK (including Tfb3, Kin28, and Ccl1) shifted TSSs upstream to the location that is used for higher eukaryotes *in vitro* (Murakami et al., 2015). A recent study from the Murakami lab has now identified that the Tfb3 N-terminus is all that is required from TFIIK for supporting the usage of downstream TSSs *in vitro* (Yang et al., 2022), consistent with the hypothesis that Tfb3 interactions with Rpb4-Rpb7 and TFIIE support TSS scanning. Our genetics support that these interactions are important *in vivo* (see below).

Here we have screened about 21,000 colonies and identified 31 *tfb3* single mutants that conferred transcriptional and/or conditional phenotypes. Phenotypes based on genetic reporters and TSS usage at *ADH1* revealed two major classes of *tfb3* mutants. Class 1 mutants showed MPA^S phenotype and tested representative alleles showed upstream shifted TSS distribution at the *ADH1* promoter. Class 2 mutants showed His^+ phenotype and tested representative alleles showed

downstream shifted TSS distribution at the *ADH1* promoter. Furthermore, substitution distribution on Tfb3 protein structure of different mutant classes revealed distinct functions of residues involved in TSS selection and further suggested class 1 and class 2 mutants as LOF and GOF mutants of Tfb3 (see below).

First, all substitutions at zinc-binding residues (C13S, C16R, and C39R) showing MPA^S phenotype suggested class 1 as a LOF class of Tfb3. Cysteine substitutions in a typical RING finger domain are expected to interfere with zinc binding and therefore disrupt the structure of the RING domain, which would be predicted as LOF for Tfb3. Based on this classification, namely LOF mutants show MPA^S, His^+ phenotypes would be predicted to be GOF. These *tfb3* mutant classes have the same characterization as *ssl2* mutant classes where structural disruptions or mutations within key residues in active sites confer MPA^S (LOF) and reduce putative TFIIH processivity, while rarer, more specific substitutions alter function or regulation as to gain activity and increase putative TFIIH processivity.

Second, substitutions located at Tfb3-Rpb7 and Tfb3-Tfa1 interfaces resulting in LOF mutants revealed the importance of Tfb3 bridging TFIIH, Pol II, and TFIIE during TSS selection. Structural studies indicate that Tfb3 bridges TFIIH, Pol II, and TFIIE via three interfaces (**Figure 8**): a charged interface formed by the Tfb3 RING finger and Pol II Rpb4-Rpb7 stalk subcomplex, a hydrophobic pocket formed by the Tfb3 RING finger, TFIIE subunit Tfa1 E-linker helices, and the Pol II stalk subunit Rpb7 OB domain, and an interface between the Tfb3 ARCH anchor and TFIIH core subunit Rad3 ARCH domain. We found two substitutions (R64K, N66D) in the Tfb3-Rpb4-Rpb7 surface and four (K17E, Y21H, L22P, and S23P) at the Tfb3-Tfa1 interface that all showed MPA^S phenotypes, predicted to disrupt interactions and result in decreased processivity of scanning. This is consistent with the hypothesis that Tfb3 tethers TFIIH to the rest of the PIC to

support a long-distance scanning process, and thus interference of surfaces is expected to decrease scanning processivity.

Third, substitutions causing GOF located at residues internal of the RING finger domain might provide some sights of how Tfb3 regulates TFIIH attachment/detachment. For example, *tfb3* F28L, I38F, and I46T alleles showed His^+ phenotypes and are predicted to be GOF mutants. All three residues are involved in a network of hydrophobic residues, which are highly conserved from humans to yeast, therefore are predicted to stabilize the RING finger core (Gervais et al., 2001). We speculate these alleles may have effects on Tfb3 conformational change delaying Tfb3 uncoupling from Rpb4-Rpb7-TFIIE, resulting in a slower dissociation rate and therefore an increase in scanning distance.

How Tfb3 communicates with other PIC components is a key open question related to Tfb3 function in transcription initiation. Studies from our lab have proposed two functional networks contributing to TSS selection. The efficiency network (Pol II, TFIIB, and TFIIF) contributes to TSS selection via efficiency of initiation and Pol II catalytic rate, whereas the processivity network (Ssl2 and Sub1) contributes to TSS selection via control of how far TFIIH can scan at a promoter. We have shown that *tfb3* mutant classes have the same characterization as *ssl2* mutant classes, where GOF mutants confer MPA^S and shift TSSs upstream while LOF mutants confer His^+ and shift TSSs downstream. Moreover, most of *tfb3* LOF substitutions are at interfaces with Rpb4-Rpb7-TFIIE and are therefore predicted to disrupt interactions and cause decreased scanning processivity, whereas GOF substitutions are mainly internal of the RING finger domain and are predicted to potentially delay TFIIH uncoupling from Rpb4-Rpb7-TFIIE via effects on Tfb3 conformational change. These features suggested that Tfb3 likely functions through the processivity network, which is further supported by primary additive genetic interactions observed

between *tfb3* alleles and *ssl2* alleles or *sub1*Δ. We have formulated the idea of two distinct networks to explain why we observe additive interactions or epistatic interactions between different factors. The interactions we observe are those predicted if some factors control initiation at all sites while others control only the probability that a site might be reached.

One subtlety in quantitative analysis of *tfb3* upstream TSS shifting mutants suggested that these *tfb3* alleles may also have increased initiation efficiency. Previously examined Pol II GOF mutants and *ssl2* LOF shifted TSS distributions upstream but in district ways. Pol II upstream shifting alleles shifted TSS distribution upstream by increasing the usage of upstream minor TSSs (poorly used TSSs), such as TSSs in the bin-2 of the *ADH1* promoter, which is predicted if the efficiency of TSS usage is affected at all TSSs. Distinctly, *ssl2* upstream shifters shifted TSS distribution upstream by decreasing the usage of downstream TSSs without activating upstream minor TSSs, which is predicted if the only process affected is the ability of scanning to reach the downstream TSSs. Surprisingly, *tfb3* upstream shifters increased the usage of both upstream major and minor TSSs (**Figure 26B)**, indicating *tfb3* alleles may have both efficiency and processivity features. The possible involvement of Tfb3 in efficiency network accords with the observed TSS effects of *tfb3* R64K and *rpb7* D166G alleles. The proposed Tfb3 R64-Rpb7 D166 salt bridge is the most important interaction that forms the charged interface between Tfb3 and Rpb7 (Schilbach et al., 2021; Schilbach et al., 2017). Interestingly, mutations at both sides of this salt bridge, *tfb3* R64K and *rpb7* D166G, that are predicted to alter the Rpb7-Tfb3 interaction, showed consistent effects on TSS selection. Specifically, the *tfb3* R64K allele showed a strong MPA^S phenotype and was therefore predicted to shift TSS usage upstream. Consistently, the *rpb7* D166G allele showed a severe MPA^S phenotype and shifted TSS usage upstream at *ADH1* (Braberg et al., 2013). Very provocatively, the *rpb7* D166G allele shifted TSS distribution upstream via the same pattern as

tfb3 upstream shifters, meaning the *rpb7* D166G allele increased the usage of both upstream major and minor TSSs. These observations together suggested that Tfb3 might be a partial efficiency factor because it functions in TSS selection via contacting with Rpb7. This hypothesis could be tested by examining genetic interactions between these alleles and Pol II GOF or TFIIF GOF alleles, which will also provide some sights to speculate how Tfb3 could function through either Pol II or TFIIH.

Combining transcriptional phenotypes, structural distribution, and TSS usage effects of identified *tfb3* alleles and their genetic interactions with *ssl2* alleles and *sub1*Δ, we speculate Tfb3 as a partial efficiency and partial processivity factor. There are a number of ways to provide explanation for this idea. First, Tfb3 may contribute to initiation efficiency through Pol II or other GTFs and to processivity through TFIIH. Second, Tfb3 may contribute to initiation efficiency and processivity both through TFIIH but via scanning rate and scanning processivity, respectively. As predicted by the Shooting Gallery model, the probability that a TSS is used (i.e., the probability a passing target is hit) is determined by both how fast the first phosphodiester bond is catalyzed (i.e., the rate of firing) and how fast the DNA template is inserted into the Pol II active site (i.e., the rate of a target passing). That means TFIIH "rate" mutants are predicted to show the same pattern of TSS effects as Pol II efficiency mutants. Thus, Tfb3 may regulate both scanning rate and processivity of TFIIH, therefore its mutants showed changes to both initiation efficiency and processivity.

To investigate these hypotheses and further study the mechanism of Tfb3 functioning in TSS selection, several experiments need to be undertaken. First, performing genomic TSS-seq on *tfb3* mutants will determine whether polar effects on TSS usage are genome-wide and provide a much greater base of data to distinguish *tfb3* alleles from other initiation mutants that have already

been analyzed. Second, investigating the genetic interactions between *tfb3* alleles and Pol II or other GTFs will provide a genetic test for additive interactions in TSS usage that would support *tfb3* mutant alterations to initiation efficiency. Third, further mutagenesis work on important residues within Tfb3-Rpb4-Rpb7 and Tfb3-TFIIE interfaces will be necessary to refine the importance of Tfb3 anchoring TFIIH to the rest of the PIC therefore supporting a long-distance scanning during TSS selection.

3.4 Materials and methods

3.4.1 Yeast strains

Yeast strains used in this chapter for *tfb3* mutant screen, *tfb3/ssl2* double mutants, *tfb3/sub1*Δ double mutants, and related experiments are listed in **Appendix B** as **Appendix Table 4-6**.

tfb3/ssl2 double mutants were constructed by two-step integration to introduce *ssl2* mutants into chromosome and plasmid shuffling to introduce *TFB3* or *tfb3* mutants on plasmid. Briefly, in step-1 of the two-step integration of *ssl2* mutants, *ssl2* mutants were introduced by chromosomal integration into *SSL2* native locus in yeast strains (CKY2212 and CKY2214) containing a chromosomal deletion of *TFB3* but with a wild type *TFB3 LEU2* plasmid (pCK1664). In step-2, duplicated target region led to loop out via integration homologous recombination to remove the plasmid sequence. To introduce *tfb3* mutants, the *TFB3* WT *LEU2* plasmid was first replaced with a *TFB3* WT *URA3* plasmid (pCK1632) by SC-Ura plate selection. Next, *LEU2* plasmids

containing either *TFB3* WT (as s control) or *tfb3* mutants were introduced by transformation, followed by plasmid shuffling to select strains that lost the *URA3* plasmid (Boeke et al., 1987).

*tfb3/sub1*Δ double mutants were constructed by deleting *SUB1* gene in *TFB3* shuffle strain (CKY2191 and CKY2196) genome, followed by introducing *tfb3* mutants by plasmid shuffling (Boeke et al., 1987).

All strains with deletion of *TFB3* or *SUB1* at chromosomal loci (*tfb3*Δ*::hphMX* and *sub1*Δ*::natMX*) were verified by drug resistance to corresponding antibiotic section markers (hygromycin B for *tfb3*Δ*::hphMX* and nourseothricin sulfate for *sub1*Δ*::natMX*), PCR genotyping, and sequencing. All strains with mutations at *SSL2* chromosomal loci were verified by PCR genotyping and sequencing.

3.4.2 Plasmids

Plasmids used in this chapter for *tfb3* mutant screen, *tfb3/ssl2* double mutants, *tfb3/sub1*Δ double mutants, and related experiments are listed in **Appendix B** as **Appendix Table 8-10**.

3.4.3 Oligonucleotides

Oligonucleotides used in this chapter for *tfb3* mutant screen, *tfb3/ssl2* double mutants, *tfb3/sub1*Δ double mutants, and related experiments are listed in **Appendix B** as **Appendix Table 12-14**. All oligonucleotides were obtained from IDT.

3.4.4 Yeast media

Yeast media were prepared following standard protocols (Amberg et al., 2005). YP solid medium is made of yeast extract (1% w/v; BD), peptone (2% w/v; BD, 211677) and bacto-agar (2% w/v; BD, 214010), supplemented with adenine (0.15 mM; Sigma-Aldrich, A9126) and L-tryptophan (0.4 mM; Sigma-Aldrich T0254). YPD plates contained dextrose (2% w/v; VWR, VWRBK876), YPRaf plates contained raffinose (2% w/v; Amresco, J392), and YPRaf/Gal plates contained raffinose (2% w/v; Amresco, J392) and galactose (1% w/v; Amresco, 0637) as carbon sources. YPRaf and YPRaf/Gal plates also contained antimycin A (1 μg/ml; Sigma, A8674-100mg). Minimal media plates are synthetic complete ("SC") with amino-acids dropped out as appropriate as described in (Amberg et al., 2005) with minor alterations as described in (Kaplan et al., 2012): per standard batch formulation, adenine hemisulfate (Sigma-Aldrich, A9126) was 2 g, L-Leucine (Sigma-Aldrich, L8000) was 4 g, myo-inositol was 0.1 g, para-aminobenzoic acid (PABA) was 0.2 g. SC-Leu+5FOA plates contained 1 mg/ml final concentration of 5-fluoroorotic acid monohydrate (5-FOA, GoldBio, F-230). SC-Leu+MPA plates contained 20 μg/ml final concentration of mycophenolic acid (MPA, Sigma, M3536-250MG). SC-His+3AT plates contained different final concentrations of 3-Amino-1,2,4-triazole (3-AT, Sigma-Aldrich, A8056).

3.4.5 Genetic screening for transcription-related phenotypes

Mutagenic PCR and gapped plasmid preparation. To generate randomly mutagenized *TFB3*, mutagenic PCR was performed as six parallel Taq (NEB, M0267L) PCR reactions with eight repeats for each, using pRSII316-*TFB3* (pCK1632) as template and CKO1800 and CKO1799 as oligos. 25 cycles were used to minimize multiple mutations. Gel purification using E.Z.N.A.®

Gel Extraction Kit (Omega Bio-Tek, D2500-02) and ethanol precipitation (0.1 volume of 3M sodium acetate and 3 volumes of 100% ethanol) were performed to clean up and concentrate PCR products.

To generate gapped transformation plasmid, pRS315-*TFB3* (PCK1664) plasmid was double digested by FastDigest enzymes BglII (Thermo Scientific, FD0083) and NruI (Thermo Scientific, FD2154), following by gel purification and ethanol precipitation.

Gap repair transformation into yeast and transcription-related phenotypes screening. Yeast strains with *tfb3*Δ under *imd2*Δ*::HIS3* (CKY2191) or *IMD2* (CKY2196) background were streaked on YPD plates for 2 days. Single colonies were inoculated into 5 ml YPD media for overnight. On the day of transformation, 1 ml saturated overnight culture was inoculated into 50 ml YPD media and grown at 200 rpm until $2x10^7$ cell/ml, as determined by cell counting. 10 µg mutagenic PCR products and 500 ng gapped plasmid, resulting in a molar ratio of insert and vector of ~ 100:1, were transformed following yeast high-efficiency transformation protocol described in (Gietz and Schiestl, 2007). To check transformation efficiency and therefore limit colony density on screening plates to 300-400 colonies per plate, different amounts of transformants were plated on SC-Leu plates for counting, and the remaining transformation reactions were stored at 4°C. Next, appropriate amounts of transformation reactions were plated on 20 SC-Leu plates per strain, taking into account empirically determined loss in viability of transformation mixes stored at 4˚C. After three days of growth, screening plates were replicated onto SC-Leu+5FOA plates to select cells that had lost the *TFB3 URA3* plasmid. After four days subsequent growth, SC-Leu+5FOA plates were replicated onto phenotyping plates, including SC-Leu+20 µg/ml MPA specifically for *IMD2* background, SC-His+1 mM 3AT plate for *imd2*Δ*::HIS3* background, and SC-Leu, SC-Lys, YPD (at 30°C, 37°C, 16°C), YPRaf and YPRaf/Gal plates for both backgrounds. Phenotypes were

recorded over seven days of subsequent growth. Candidates were streaked from SC-Leu phenotyping plates onto Sc-Leu plates, followed by replicating onto all phenotyping plates to confirm phenotypes.

Yeast plasmid rescue and sequencing. Yeast plasmid DNA was isolated using ZR Plasmid miniprep - Classic (ZYMO Research, D4016) per manufacturer′s instructions, with 0.5 mm glass beads added after adding the cell resuspension solution (P1) for breaking cells. Next, plasmids were transformed into homemade *Escherichia coli* TOP10F′ cells and grown on LB+Carb plates. Representative colonies were picked and streaked on LB+Carb plates for singles. Plasmids were isolated using ZR Plasmid miniprep - Classic (ZYMO Research, D4016) and stored. Plasmids containing mutant candidates were sent for sequencing to identify mutation(s) within *TFB3*.

Phenotyping via plasmid shuffle and spot assays. To confirm plasmid linkage with observed phenotypes, plasmids extracted from *E. coli* were transformed into the two *TFB3* shuffle strain backgrounds (CKY2191 and CKY2196) following yeast regular efficiency transformation protocol. Next, representative colonies were picked and patched on SC-Leu plates. After fully grown, plates were replicated to SC-Leu+5FOA to select yeast cells with wild-type *TFB3 URA3* plasmid shuffled out and with *tfb3* mutant *LEU2* plasmid retained. Growth phenotypes were recorded over seven days. After patch assay, either spot phenotyping assays or patch phenotyping by replica plating were performed.

To perform spot assays, two candidates per mutant or WT being tested were selected from representative patches and streaked on SC-Leu+5FOA plates for single colonies. Next, single colonies were inoculated into 2 ml YPD media for growth until saturation. Subsequentially, 5 x 10-fold serial dilution in H_2O was performed for each saturated culture and then applied to phenotyping plates using a 48-well metal pinner. Plates included SC-Leu, SC-Leu+20 µg/ml MPA,

SC-Lys, SC-His+1 mM/2 mM/5 mM/10 mM 3AT, YPD (at 30°C, 37°C, 16°C), YPRaf and YPRaf/Gal plates, for phenotype test. Phenotypes were documented over seven days.

To perform patch phenotyping assay for primary investigation of MPA^S and His+ phenotypes, patch plates were replicated onto phenotyping plates, including SC-Leu, SC-Leu+20 µg/ml MPA, and SC-His+1 mM/10 mM 3AT plates. Phenotypes were documented over four days.

3.4.6 Primer extension

Total RNA was extracted by a phenol-chloroform method (Schmitt et al., 1990). Primer extension assays were performed as described in (Ranish and Hahn, 1991) with modifications described in (Kaplan et al., 2012). For each reaction, 30 µg total RNA was used. An RNA sample from a *TFB3* strain was used as "WT" control. A sample of nuclease-free water was used as "no RNA" control. A primer (CKO401) complementary to *ADH1* mRNA was labeled with ^{32}P γ-ATP (PerkinElmer, BLU502Z250UC) and T4 polynucleotide kinase (Thermo Scientific, EK0031). M-MuLV Reverse Transcriptase (NEB, M0253L), RNase inhibitor (NEB, M0307L), dNTP, and DTT were added to RNA and labelled primer mix for reverse transcription reaction. Before loading to sequencing gel, RNase A (Thermo Scientific, EN0531) was added to remove RNA. The products were analyzed by 8% acrylamide/bis-acrylamide (19:1 ratio, Bio-Rad, 1610145) gel containing 1x TBE and 7 M urea. Primer extension gel was visualized by Molecular Imager PharosFX™ Plus System (Bio-Rad) and quantified by Image Lab (5.2).

4.0 Summary and future directions

This dissertation studied mechanisms of how promoter DNA sequence, Pol II catalytic activity, and the TFIIH subunit Tfb3 determine TSS selection by promoter scanning in *Saccharomyces cerevisiae*. In **Chapter 2**, I determined DNA sequence and Pol II activity contributions to initiation efficiency in a controlled context using a developed "Pol II MASTER" system. In **Chapter 3**, I explored Tfb3 effects on TSS selection using genetic and molecular approaches.

What sequences make efficient TSSs? To dissect Pol II TSS sequence specificity and how Pol II activity alters it, we developed a massively parallel reporter assay Pol II MASTER and examined TSS efficiencies for ~1 million unique TSS sequences within a controlled context. We find that sequences at individual positions at or upstream of TSSs modulate initiation efficiency over a wide range and in a predictable way. In addition, we find that sequences downstream of TSS also contribute to TSS efficiency. However, limited by the original design of our libraries, the understanding of these TSS downstream positions and their potential interactions with upstream positions need to be specifically studied at a higher resolution. Therefore, we plan to construct massive promoter variant libraries with a larger randomized region at TSS downstream and technical additions to allow precise measurement of overall expression, to get a comprehensive understanding of DNA sequence contributions to TSS selection and expression level. We find functional interactions between neighboring positions, suggesting they might function together. Combining results from our Pol II MASTER study and others, we proposed that two major groups of positions contribute to TSS selection: bases around TSS (actual initiating site) and bases around position -8. First, the TSS and adjacent bases interact with the Pol II active site, the 1st NTP, or

each other to stabilize or facilitate the first few NTPs binding and therefore stimulate RNA synthesis. Second, we speculate that Ts around position -8 on the template strand (As on the coding strand) function together with TFIIB to pause the scanning process and therefore hold potential TSSs positioned about 8 bases downstream in the Pol II active site longer to facilitate initiation. This proposed mechanism suggests a considerable role of TFIIB in TSS selection. Therefore, we plan to introduce our Pol II MASTER into strains with TFIIB alleles that have been shown or predicted to have effects on TSS selection to directly test the proposed TFIIB function in TSS selection via interaction with DNA sequence.

How does altered Pol II catalytic activity interact with initiation? Pol II catalytic mutants previously have been observed to alter TSS usage of motifs in the genome. However, because of confounding factors in the genome, we were not able to distinguish whether it was a direct or an indirect effect of altered Pol II activity or more apparent than real due to sequence biases within genomic promoters. To investigate how Pol II activity alters TSS sequence specificity, we examined Pol II mutant effects on TSS preference by comparing TSS efficiencies of TSS variants in our controlled promoter libraries in Pol II WT and mutants. We find that Pol II mutants directionally altered TSS efficiency across all TSS motifs. Specifically, Pol II hyperactive mutants increased overall efficiency for all TSS sequences, while Pol II hypoactive mutants decreased overall efficiency for all TSS sequences. This revealed that previously observed apparent alteration of A_{-8} and B_{-8} preference of Pol II mutants at genomic promoters was likely an indirect effect due to sequence skew. The apparent changes observed were caused by shifted TSS distribution and the uneven distribution of TSS motifs within promoters. Moreover, no selective alteration in preference for A_{-8} motifs also suggested that Pol II catalytic activity does not specifically interact with sequence at position -8. We did observe selective effects of Pol II mutants for the +1 position

(actual initiation site). This could be interpreted as a direct effect of Pol II mutants on catalysis for ATP vs GTP substrates. Alternatively, it could be an indirect effect of a potentially altered GTP/ATP ratio in Pol II mutant strains. To investigate this latter hypothesis, different types of experiments will be undertaken. First, to investigate whether and how the GTP/ATP ratio was changed by Pol II mutants, we will measure concentrations of ATP and GTP in WT and Pol II mutants. Second, to directly examine whether and how different NTP conditions affect Pol II initiation sequence preference, we will collaborate with Kenji Murakami's lab (University of Pennsylvania) and perform transcription initiation on our Pol II MASTER libraries using their *in vitro* reconstituted system over a range of NTP concentrations, including alteration of the ATP/GTP ratio. Third, to examine how altered cellular GTP concentration affects TSS selection, we will treat WT cells containing our Pol II MASTER libraries with the drug MPA (Mycophenolic Acid), which depletes GTP in cells. This experiment will also give us an opportunity to extend our understanding of how cellular state, such as altered host metabolism, contributes to TSS selection.

To what extent does DNA sequence around a TSS contribute to TSS efficiency at genomic promoters? We used regression modeling to identify robust DNA sequence features (key base identifies and their interactions) and compared the difference between observed and model-predicted efficiencies of positions within known genomic promoter windows. We observed that most promoter positions showed low or no observed efficiency and were over-predicted from sequence alone. This is expected because TSSs need to be specified in an appropriate chromatin context and also suggested contributions of additional attributes on TSS efficiency beyond local sequence. Therefore, to determine contributions of other promoter architectural factors on Pol II initiation, we have designed and are constructing "architecture" libraries that apply developed Pol II MASTER analysis to other Pol II initiation regulatory elements, such as core promoter-TSS

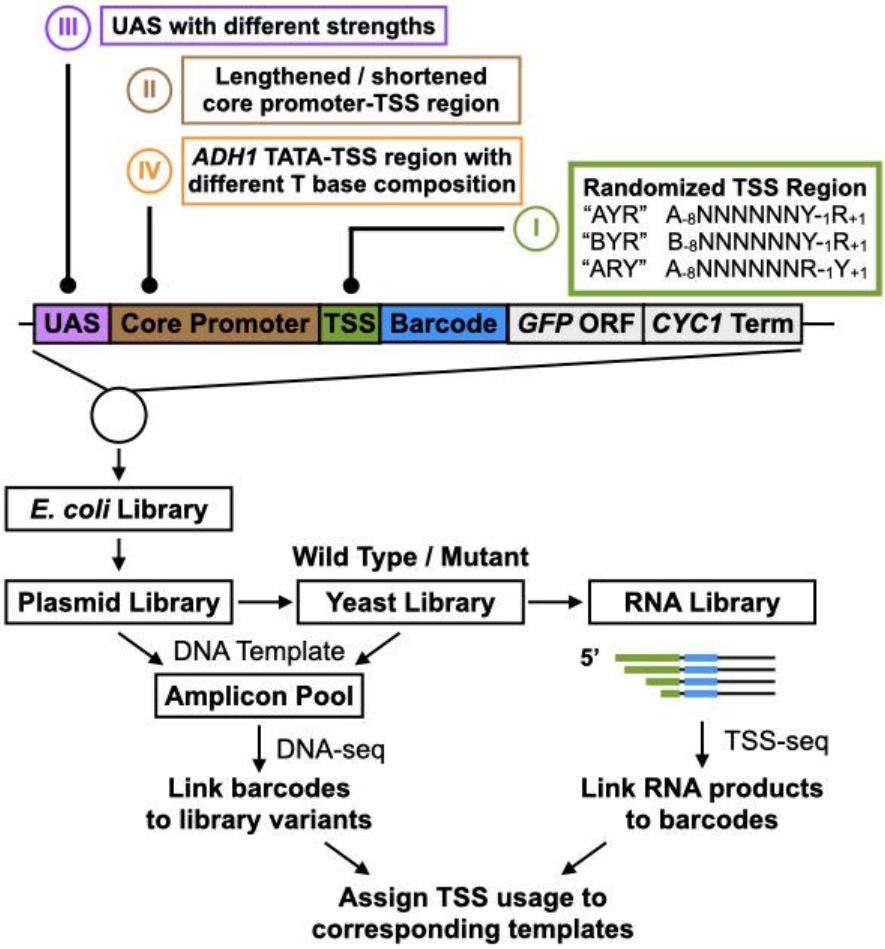
distance, UAS identity, and sequence composition within the scanning region. Details of these architecture libraries will be discussed in **Appendix A**. Our final goal is to quantitatively model and predict Pol II initiation output for any particular promoter. After we individually determine features of promoter architectural factors, we will train a deep learning model by combining data from different libraries.

How does Tfb3 participate in transcription initiation? We successfully identified two classes of *tfb3* mutants, namely mutants that shift TSS distribution upstream or downstream, using genetic reporters. Both classes have been validated by examining TSS usages at the *ADH1* promoter, but we have yet to verify these effects genome-wide. To gain an understanding of the contribution of Tfb3 on TSS selection genome-wide, we will perform TSS-seq on two classes of *tfb3* alleles to investigate global effects on TSS usage of *tfb3* alleles. *tfb3* mutant classes have the same characterization as *ssl2* mutant classes, namely LOF mutants confer MPA^S and shift TSS distribution upstream while GOF mutants confer His^+ and shift TSS distribution downstream. Moreover, broad additive genetic interactions have been observed between *tfb3* alleles and *ssl2* alleles or *sub1*Δ, both of which have been suggested to be involved in the scanning processivity network. These observations suggested Tfb3 likely functions through the processivity network. Interestingly and very provocatively, *tfb3* upstream shifters altered TSS usage in an intermediate way between known initiation efficiency and processivity mutants. Specifically, *tfb3* alleles shifted TSS usage upstream by increasing usages of both upstream minor and major TSSs, which combined features of tested efficiency mutants (Pol II alleles, increased usage of minor TSSs) and processivity mutants (*ssl2* alleles, increased usage of major TSSs). Therefore, we speculate Tfb3 could be a partial efficiency and partial processivity factor. To get additional evidence for how Tfb3 might be involved in the efficiency network, we will investigate genetic interactions of *tfb3*

alleles with initiation factors involved in the efficiency network (Pol II, TFIIB, and TFIIF). Tfb3 bridges TFIIH to the rest of the PIC via multiple interfaces with Pol II Rpb4-Rpb7 and TFIIE. These conserved contacts in the human PIC have been observed to be broken when MAT1/Tfb3 is detached from Rpb4-Rpb7 and TFIIE during initiation, which can have consequences on TFIIH detachment from the rest of the PIC. We therefore propose that Tfb3 in yeast may function similarly as its conserved human homolog MAT1, namely contributing to attachment/detachment of TFIIH, but this change may happen later during initiation in yeast, resulting in a longer scanning distance. In other words, Tfb3 has been proposed to determine the TSS usage window. Supporting this hypothesis, we have shown that mutations within proposed interfaces conferred upstream TSS shifting, consistent with the idea that these interfaces are required for normal promoter scanning. To directly examine this hypothesis and the importance of those interactions during scanning, we will perform mutagenetic studies on important residues within interfaces between Tfb3 and Rpb4-Rpb7 or TFIIE and investigate how these substitutions alter TSS usages.

Appendix A Libraries determining promoter architectural modulation of Pol II initiation

As previously described in **Section 1.2**, the scanning model predicts that the initiation efficiency of any given promoter is controlled by multiple architectural features (**Figure 1**), such as TSS sequence, core promoter-TSS distance, UAS identity, and sequence composition within the scanning region. My study of Pol II TSS sequence specificity using the developed "Pol II MASTER" approach (**Chapter 2**) emphasized the strength of our approach to minimize confounding factors by isolating specific promoter attributes. Therefore, I applied this developed systematic analysis to other regulatory elements determining Pol II initiation (**Appendix Figure 1**). The strategy is the same as that for libraries studying sequence specificity. Briefly, when studying a particular regulatory element, all other elements are kept the same for all promoter variants. After constructing libraries on plasmids and amplifying in *E. coli*, plasmid libraries are transformed into WT yeast strains to specifically characterize how a particular architectural feature contributes to TSS selection. Meanwhile, the variant libraries can be transformed into initiation factor mutants to determine how a particular initiation factor interacts with the examined architectural feature. Subsequentially, DNA-seq and TSS-seq are performed and analyzed, respectively, to get DNA template information and the TSS usage distribution for individual promoter variants. I will describe the design for each "architecture" library first and then show preliminary construction results.



Appendix Figure 1 Systematic analysis of promoter architectural effects on Pol II intiaiton

(I) Pol II "TSS" libraries constructed and analyzed in **Chapter 2** contain promoter variants with randomized positions in a specific TSS region. (II) Core promoter-TSS "distance" libraries contain *ADH1* (TATA-containing) and *RPS5* (TATA-less) promoter variants with shortened or lengthened distances between the core promoters and TSS. (III) Pol II "flux" library contains promoter variants with different expression levels and TATA-TSS distance, driven by UASs with differing TATA classes and strengths. (IV) Scanning region "sequence composition" library contains *ADH1* promoter variants with differing base composition or sequence order within the scanning region.

Appendix A.1 Core promoter-TSS distance libraries

Appendix A.1.1 Background

Previous studies have suggested limitations of the distance between core promoter – the PIC assembly position - and TSS (Faitar et al., 2001; Hahn S, 1985; Nagawa and Fink, 1985; Qiu et al., 2020; Zhang and Dietrich, 2005). Two types of distance windows have been predicted: "scanning window" in which TSSs are reachable for Pol II and "efficiency window" in which TSSs can be efficiently used (**Figure 3**). The "scanning window" is likely determined by the minimal distance required for open complex formation of the PIC (Kostrewa et al., 2009) and the maximal scanning processivity of the TFIIH subunit Ssl2 (Fazal et al., 2015). Potential TSSs that are located outside of the "scanning window" will not be able to be used. In addition, the efficiencies of TSSs closed to scanning window boundaries might be affected by distance restriction or processivity to some degree. Therefore, an "efficiency window" that supports inside TSSs to be efficiently used has been predicted as well (**Figure 3**).

Appendix A.1.2 Rationale

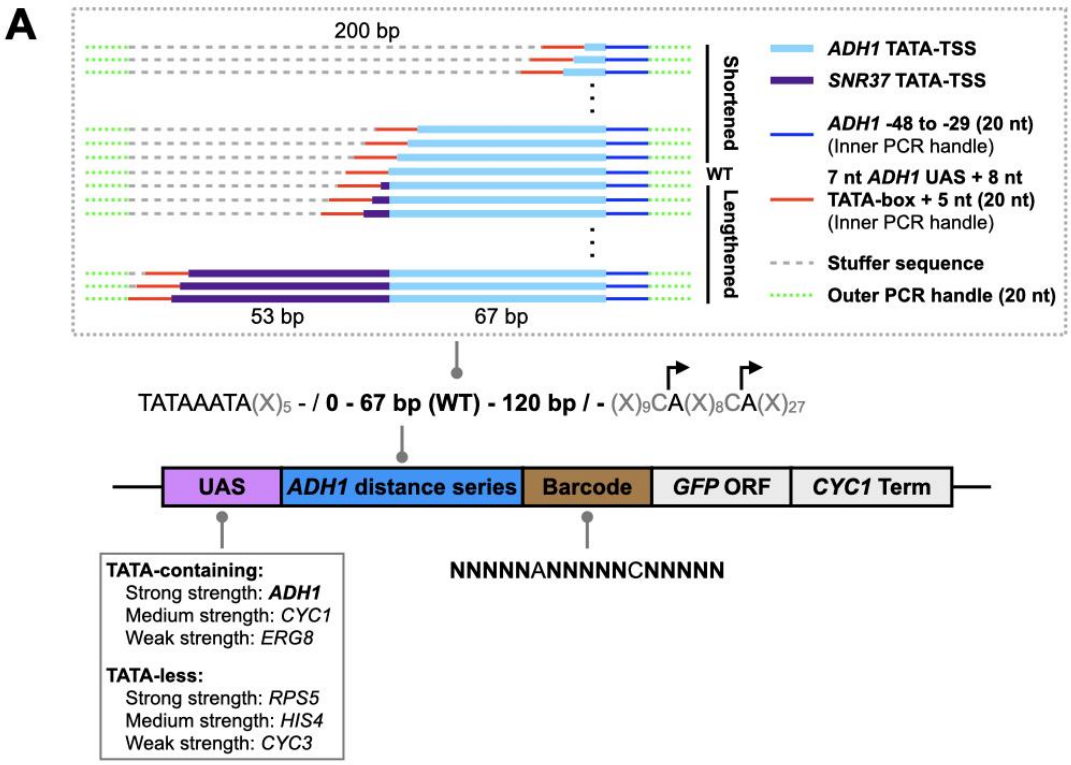
To determine how core promoter-TSS distance controls Pol II initiation output and how different promoter identities affect promoter distance constraints, I have designed and preliminarily constructed a series of *ADH1* (TATA-containing) or *RPS5* (TATA-less) promoter variants with differing core promoter-TSS region distances. Multiple TSSs are located within the TSS region. When the TSS region gets too close to the core promoter, the usage of upstream TSS(s) within the TSS region is expected to shift to downstream TSS(s). Similarly, as a TSS region being

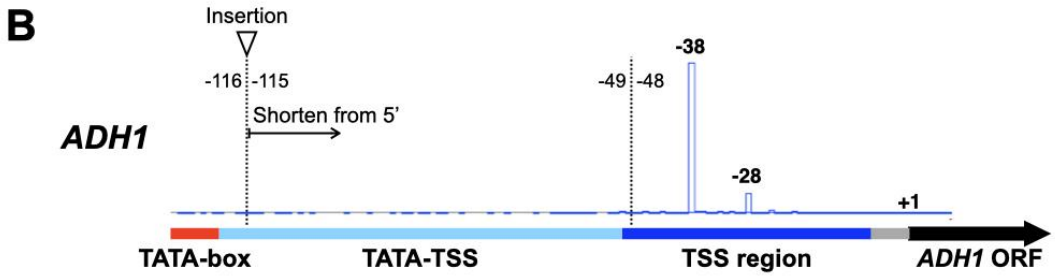
moved away, the usage of downstream TSS(s) within the TSS region is expected to get worse. From these libraries, we will determine the "efficiency window" in which TSSs can be efficiently used and the "scanning window", namely the initiation boundary, where initiation efficiency goes to zero. We will also determine whether different classes of promoters (TATA-containing versus TATA-less promoters) have similar distance constraints. Further, we can test promoter identity effects on distance constraints as well. We will compare UASs with differing strengths driving the same core promoter-TSS region to ask if they have the same TSS usage distributions and efficiencies.

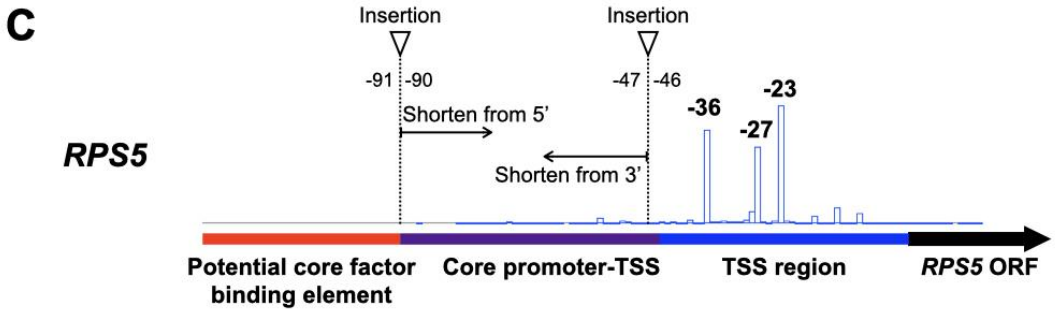
Appendix A.1.3 Experimental design and interpretation

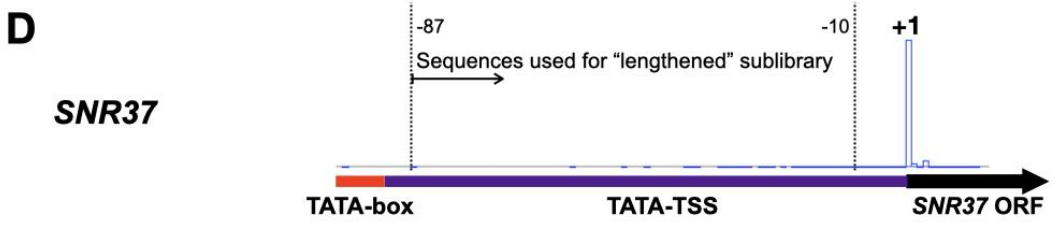
A series of pre-barcoded *ADH1* promoter variants (121 variants in total) with shortened or lengthened distances between TATA box and TSS were synthesized and are being introduced into plasmid backbones (**Appendix Figure 2**). The *ADH1* promoter contains one TATA element (TATAAATA) located between positions -128 to -121 (the translation-initiating ATG position designated as +1) with multiple TSSs distributed over a range of positions between -50 to -7, with two major TSSs (-38 and -28 TSSs) (**Appendix Figure 2B**) (Pinto et al., 1992; Qiu et al., 2020). In a previous primer extension analysis, 40 bp deletion of TATA-TSS region within *ADH1* promoter caused a significant decrease of the usage of the first major TSS (-38 TSS) but concomitant increase in the usage of minor downstream starts, suggesting that this distance (43 bp) is close to the upstream initiation boundary of the *ADH1* promoter (Faitar et al., 2001). For *ADH1* "shortened" variants, we shortened the distance over a greater range to ask when the second major TSS (-28 TSS) will be affected and when the initiation efficiency will be completely eliminated (**Appendix Figure 2A, B**). Toward achievement of this goal, sequences between

positions -115 to -49 of *ADH1* promoter were shortened from 5′ until the entire region (67 bp) is deleted. For *ADH1* "lengthened" variants, sequences from the TATA-TSS region (scanning region) of *SNR37* with varying lengths ranging from 1 bp to 53 bp were inserted between positions -116 and -115 of the *ADH1* promoter (**Appendix Figure 2A, B, D**). *SNR37* contains a single and ultra-strong TSS (**Appendix Figure 2D**) without TSS usage or preferred $A_{-8}Y_{-1}R_{+1}$ motif observed in its scanning region (Qiu et al., 2020), therefore no TSS is expected to be introduced to the inserted region of "lengthened" variants. Furthermore, the series of *ADH1* promoter variants can be paired with different UASs from different promoter classes and with different strengths (**Appendix Figure 2A**) to investigate whether or how promoter identity interacts with the distance constraint.


A
200 bp
Shortened
WT
Lengthened
53 bp
67 bp
ADH1 TATA-TSS
SNR37 TATA-TSS
ADH1 -48 to -29 (20 nt)
(Inner PCR handle)
7 nt ADH1 UAS + 8 nt
TATA-box + 5 nt (20 nt)
(Inner PCR handle)
Stuffer sequence
Outer PCR handle (20 nt)
TATAAATA(X)5 - / 0 - 67 bp (WT) - 120 bp / - (X)9CA(X)8CA(X)27
UAS
ADH1 distance series
Barcode
GFP ORF
CYC1 Term
TATA-containing:
Strong strength: ADH1
Medium strength: CYC1
Weak strength: ERG8
TATA-less:
Strong strength: RPS5
Medium strength: HIS4
Weak strength: CYC3
NNNNNANNNNNCNNNNN


B
ADH1
Insertion
-116 -115
Shorten from 5'
-49 -48
-38
-28
+1
TATA-box
TATA-TSS
TSS region
ADH1 ORF


C
RPS5
Insertion
-91 -90
Shorten from 5'
Insertion
-47 -46
Shorten from 3'
-36
-27
-23
Potential core factor
binding element
Core promoter-TSS
TSS region
RPS5 ORF


D
SNR37
-87
Sequences used for "lengthened" sublibrary
-10
+1
TATA-box
TATA-TSS
SNR37 ORF

Appendix Figure 2 Design of *ADH1* and *RPS5* core promoter-TSS distance libraris

(A) Design of *ADH1* promoter-TSS distance library. Upper panel within the dashed box shows the composition of 200 nt oligo pool of *ADH1* promoter-TSS distance variants. Two outer PCR handles (in dashed green) were used to amplify variants belonging to this particular library from a larger and mixed synthesized oligo pool. Two inner handles (in red and dark blue) were used to further amplify library variants without "stuffer sequences" (in dashed gray), which are sequences from *YDR222W* flanking region and are required to make sure all oligos have the same 200 nt length during pool synthesis. Lower panel shows compositions of *ADH1* distance library, including the *ADH1* UAS, *ADH1* distance series, a barcode region, the *GFP* ORF, and the *CYC1* terminator. The 17 nt randomized barcode region, with 2 fixed bases inserted between every 5 nt, allows us to link RNA products to corresponding DNA templates. The initial *ADH1* distance library uses native *ADH1* UAS and contains in total 121 variants. In the future, different UASs can be tested. **(B)** TSS usage distribution of the *ADH1* promoter (Qiu et al., 2020) and strategies of deletion and insertion. *ADH1* "shortened" variants contain varying *ADH1* TATA-TSS regions, where shortening starts from position -115 to -49. *ADH1* "lengthened" variants contain an insertion of *SNR37* sequences with varying lengths (details in **D**) between positions -116 and -115. **(C)** TSS usage distribution of the *RPS5* promoter (Qiu et al., 2020) and strategies of deletion and insertion. *RPS5* "shortened" variants contain varying *RPS5* core promoter-TSS regions, where shortening starts from position -115 to -49 or from position -49 to -115. *RPS5* "lengthened" variants contain an insertion of *SNR37* sequences with varying lengths (details in **D**) between positions -91 and -90 or between positions -47 and -46. **(D)** TSS usage distribution of the *SNR37* promoter (Qiu et al., 2020) and sequences used for "lengthened" variants.

The design for *RPS5* (TATA-less) core promoter-TSS distance library (216 variants in total) is essentially the same as that for the *ADH1* distance library (**Appendix Figure 2A, C**). Instead of a consensus TATA-box as PIC assembly position at the *ADH1* promoter, the region between positions -125 to -91 relative to ATG of *RPS5* has been proposed to be the potential core factor binding element (Kamenova et al., 2014; Sugihara et al., 2011). Additionally, it is unclear whether the sequences around the potential core factor binding element have additional contribution to the scanning process, therefore both "shortened" and "lengthened" sub-libraries

have two versions: one from 5′ and one from 3′ (**Appendix Figure 2C**). Specifically, for "shortened" variants, *RPS5* core promoter-TSS region was deleted either from positions -90 to -47 or from positions -47 to -90. For "lengthened" variants, sequences of *SNR37* scanning region were inserted either between positions -91 and -90 or between positions -47 and -46.

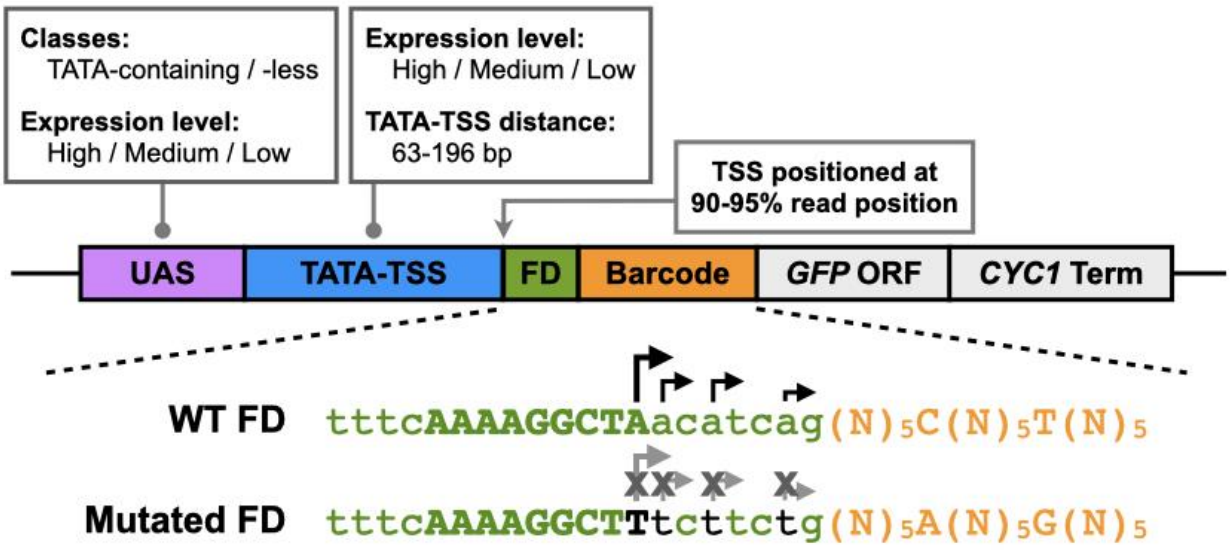
Appendix A.2 Pol II "flux" library

Appendix A.2.1 Background

A previous study has suggested that native promoters might waste some Pol II flux (the amount of Pol II recruited to a particular promoter) (Lubliner et al., 2015), which may be caused by the absence of efficient TSSs to capture flux before processivity limit is reached (**Figure 3A**). The scanning model predicts that efficient conversion of Pol II flux to initiation can limit the downstream edge of the scanning window in that while scanning *could* go further, it doesn't lead to increased initiation because all polymerases have initiated (**Figure 3C**). In addition, synthetic hybrid promoter studies have shown that UASs can positively and negatively interact with core promoters to regulate expression (Blazeck et al., 2012; Dhillon et al., 2020; Liu et al., 2020). Moreover, a UAS has been shown to be able to proximally initiate transcription without the presence of a nearly core promoter (Dobi and Winston, 2007). These observations and predictions raise two intriguing questions about promoter identity: **(1) how efficiently do genomic promoters convert recruited Pol II into initiation** and **(2) do UASs also regulate TSS distributions in addition to overall expression levels?**

Appendix A.2.2 Rationale

To determine whether Pol II flux is typically fully converted to initiation and how UASs interact with core promoters to contribute to TSS selection, I have designed and preliminarily constructed a series of TATA-containing promoter variants with different native expression levels and TATA-TSS distances, driven by UASs normally driving differing promoters from different classes and strengths, with either a WT or mutated "Flux Detector" (FD) inserted at the downstream edge of the native initiation region (**Appendix Figure 3**). The FD – the highly efficient *SNR37* TSS region – can capture the remaining Pol II flux from upstream TSS regions. For a particular UAS-core promoter variant, if its WT FD version captures additional TSS usage that is not observed in its mutated FD version, this would suggest that Pol II flux might be wasted at some promoters (**Figure 3A, B**). We will determine whether different UAS-core promoter couplings affect TSS selection by comparing the shapes of TSS distribution of a particular core promoter driven by different UASs (**Figure 3C, D**). We will also investigate how promoter class and expression level of UAS interact with expression level and TATA-TSS distance of core promoter by examining different couplings. This result will also provide useful information about UAS choices for studying how UAS affects core promoter-TSS distance constraint proposed in distance libraries (**Appendix A.1**).



Appendix Figure 3 Design of Pol II flux library

UASs with differing TATA classes and strengths are coupled core promoters with different expression levels and TATA-TSS distances. A WT or mutated FD was placed right after the native TSS that is located between 90th-95th percentile read positions of TSS-seq data from our lab (Qiu et al., 2020). The WT FD is 20 nt of the native *SNR37* TSS region and the same as the FD used in my sequence specificity libraries (**Chapter 2**). Compared to WT FD, the mutated FD has four As that have been observed TSS usage in my sequence specificity library being mutated to Ts to kill their initiation potential. The 17 nt randomized barcode region, with 2 fixed bases inserted between every 5 nt, allows us to link RNA products to corresponding DNA templates.

Appendix A.2.3 Experimental design and interpretation

A series of pre-barcoded promoter variants with different expression levels and TATA-TSS distances, driven by UASs with differing TATA classes and strengths, has been designed and is being constructed (**Appendix Figure 3**). In total, 6 relatively well studied UASs (*ADH1*, *CYC1*, *ERG8*, *RPS5*, *HIS4*, and *CYC3*) (**Appendix Table 1**) and 36 promoters (TATA-TSS region) (**Appendix Table 2**) have been selected for coupling based on their promoter properties, resulting in 432 promoter variants in total. The expression levels of native promoters are based on TSS-seq

data from our lab (Qiu et al., 2020) and subgrouped by “low expression”, “medium expression” and “high expression” (defined in **Appendix Table 2**). The downstream edge of the TATA-TSS region is the TSS that is located between 90^{th}-95^{th} percentile read position of the examined promoter, termed “selected TSS”.

Appendix Table 1 Promoter properties of selected UASs

Expression levels (Exp.) of promoters are defined based on their total TSS-seq reads within known promoter windows and grouped: low, [0, 1000); medium, [1000, 5000); high, [5000, max).

Standard name	**Systematic name**	**TATA-class**	**Exp. (reads)**	**UAS region (ATG as +1)**	**Ref.**
ADH1	YOL086C	TATA-containing	1,286,755 (high)	-414 to -129	(Ruohonen et al., 1991)
CYC1	YJR048W	TATA-containing	3,869 (medium)	-615 to -228	(Guarente et al., 1984; Guarente and Mason, 1983)
ERG8	YMR220W	TATA-containing	883 (low)	-846 to -81	(Tsay and Robinson, 1991)
RPS5	YJR123W	TATA-less	1,343,551 (high)	-500 to -246	(Sugihara et al., 2011)
HIS4	YCL030C	TATA-less	2,220 (medium)	-523 to -124	(Devlin et al., 1991; Donahue et al., 1983)
CYC3	YAL039C	TATA-less	466 (low)	-505 to -106	(Bernard et al., 2003; Dumont et al., 1987)

Appendix Table 2 Promoter properties of selected core promoter-TSS variants

The 36 selected core promoter-TSS variants were shown in three subgroups defined by expression level. Expression levels (Exp.) of promoters are defined based on their total TSS-seq reads within known promoter windows and grouped: low, [0, 1000); medium, [1000, 5000); high, [5000, max). The "TATA-selected TSS (bp)" represents the distance between TATA-box and the selected TSS, which is the TSS located between 90^{th}-95^{th} percentile read position of the examined promoter.

Low exp. group			**Medium exp. group**			**High exp. group**		
Systematic Name	**TATA-selected TSS (bp)**	**Exp. (reads)**	**Systematic Name**	**TATA-selected TSS (bp)**	**Exp. (reads)**	**Systematic Name**	**TATA-selected TSS (bp)**	**Exp. (reads)**
YJL213W	93	992	YGL186C	63	4,737	YOL086C	105	1,286,755
YOR107W	97	915	YBR105C	120	4,665	YBR031W	90	725,309
YHR037W	104	873	YOR344C	191	3,827	YHR143W	86	258,256
YLL028W	103	836	YDR244W	100	3,375	YDR224C	111	88,252
YIL015W	117	798	YBR126C	118	2,955	YML028W	86	75,597
YMR009W	76	797	YER056C	96	2,935	YHR008C	50	65,249
YAL005C	125	794	YPL135W	138	2,475			
YKL001C	110	790	YLR258W	111	2,219			
YJR080C	131	744	YBL003C	165	2,179			
YPR109W	122	701	YPR191W	78	2,155			
YER124C	123	691						
YDR508C	123	652						
YLR297W	85	563						
YPL053C	108	556						
YBR135W	139	535						
YNL111C	137	520						
YMR195W	196	460						
YDR195W	106	446						
YKL106W	157	440						
YMR316W	73	414						

Appendix A.3 Scanning region sequence composition library

Appendix A.3.1 Background

Previous studies have suggested that sequence composition within the scanning region, the region between the PIC assembly position (a TATA-box if present) and TSSs, contributes to initiation regulation. Specifically, pyrimidine-richness, especially T-richness on the coding strand, within scanning regions correlate with, and appears to contribute to, high expression (Lubliner et al., 2013; Lubliner et al., 2015; Maicas and Friesen, 1990). T-richness has been suggested to be able to promote nucleosome depletion because G/C content promotes nucleosome occupancy (Lee et al., 2007; Peckham et al., 2007; Tillo and Hughes, 2009). T-richness on the coding strand and corresponding A-richness on the template strand can facilitate DNA melting by providing an easily meltable region and therefore help transcription bubble formation (Bansal et al., 2014). Moreover, depending on the characteristics of the sequence itself, T-rich regulatory sequences proximal to TATA-box can define the 5′ boundary of functional transcription window by promoting Cryptic Unstable Transcripts (CUTs) degradation (McNeil, 1988; Thiebaut et al., 2008) and/or by preventing initiation from upstream promoters interfering with downstream or reading entirely through the downstream ORF. These observations emphasized the influences of T-richness on the coding strand on promoter activity. However, other base features such as G/C% have also been suggested to compete with T-richness to modulate initiation (Liu et al., 2020). This raises the need for a more comprehensive study of effects of scanning region sequence composition on initiation, including both overall expression level and TSS selection.

Appendix A.3.2 Rationale

To determine how sequence composition within the scanning region contributes to TSS selection, I have designed and preliminarily constructed a series of *ADH1* promoter variants with different T% or sequence order. *ADH1* is a highly expressed gene and has a T-rich scanning region. Among 70 bp scanning region (-120 to -51 relative to the translation-initiating ATG) of *ADH1*, there are 38 Ts and 32 non-Ts (18 Cs, 7 As, and 7 Gs) (**Appendix Figure 4**). To investigate sequence composition features using the *ADH1* promoter, two main classes of modifications have been made for the scanning region. First, the percentage of Ts has been decreased or further increased to determine how T content affects not only overall expression level but also TSS distribution. Second, the sequences of *ADH1* scanning region have been reordered to investigate how characteristics of the sequence itself contribute to initiation.

Appendix Figure 4 The T-richness of *ADH1* scanning region

Sequences of *ADH1* scanning region are shown with Ts in red. This 70 bp scanning region is highly pyrimidine-rich, especially T-rich, and contains 38 Ts, 18 Cs, 7 As, and 7 Gs. Position numbers relative to the translation-initiating ATG are given for scanning region borders.

Appendix A.3.3 Experimental design and interpretation

A series of pre-barcoded *ADH1* promoter variants with different modifications within the scanning region has been designed and is being constructed. First, to examine how decreasing T% on the coding strand within the scanning region affects initiation, different amounts of Ts, ranging

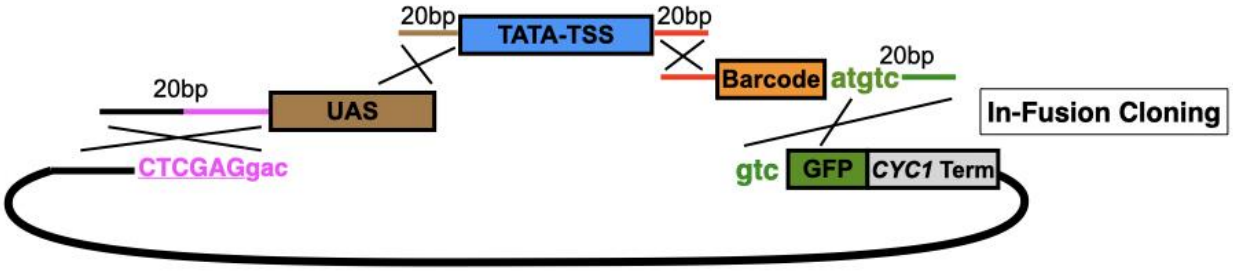
from 1 to 37, have been randomly selected from 38 Ts between positions -117 to -51 and replaced with C(s) (to maintain pyridine richness on the coding strand). For each level of replacement, meaning replacement for a certain amount of Ts, 10 different variants were chosen for test. Therefore, in total 370 promoter variants contain decreased T content compared to WT *ADH1*, designated as "T-decreased" variants. Second, to examine how increasing T% within the scanning region affects initiation, different amounts of non-Ts, ranging from 1 to 28, have been randomly selected from 29 non-Ts between positions 117 to -51 and replaced with T(s). The strategy for random selection is the same as that used for "T-decreased" variants. In total, 280 promoter variants contain increased T content, designated as "T-increased" variants. Third, to examine how sequence order or potential sequence motifs affect initiation, 67 bp sequences between positions -117 to -51 have been randomly re-ordered to generate 20 promoter variants, which comprise the same base composition, but different sequence order compared to WT *ADH1*. All promoter variants are driven by the native *ADH1* UAS and contain the same TSS region.

Appendix A.4 Preliminary construction of libraries

I have constructed 1st generation of above "architecture" libraries and developed preliminary analysis pipelines to process DNA-seq and TSS-seq results. However, analysis results suggested two issues: template switching happened during plasmid library construction and low diversity interfered with pool sequencing quality of Read 2 of TSS-seq. I will briefly describe these issues and solutions.

First, observed template switching was found to be due to the use of In-Fusion cloning in initial plasmid generation and can be solved by ligation strategy. The construction strategy of

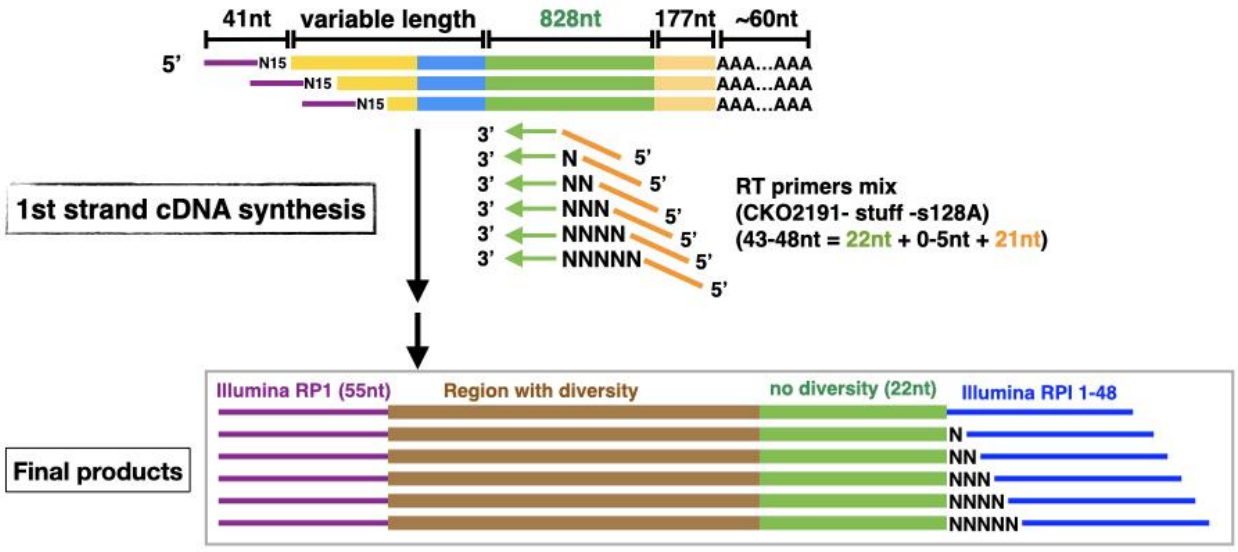
plasmid libraries was originally designed to use In-Fusion cloning system (**Appendix Figure 5** showing Pol II "flux" library as an example) because of its two major advantages. First, In-Fusion enzyme fuses DNA fragments by recognizing 15-20 bp overlaps at their ends and allows multiple fragments simultaneously cloned into vector in a single reaction. In this way, sewing PCR reactions that fuse different template components (such as UAS, TATA-TSS, and barcode) can be skipped. Additionally, it would be easier to modularly update libraries, such as adding more UAS variants, with no need to reconstruct the entire template. Second, In-Fusion cloning is ligation-independent and therefore prevents self-ligation. However, because of the nature of our libraries, many identical or highly similar sequence segments exist in a majority of promoter variants but within different contexts. Because of this, these sequence segments were incorrectly recognized by In-Fusion enzyme as overlaps, allowing substantial template switching. To solve this, I changed from the In-Fusion strategy to PCR sewing plus ligation. Because the latter strategy was originally considered as an alternative approach in library design, the majority of the library design is compatible for both approaches. To test the new strategy, I have re-constructed the transcription template fragments of *ADH1* "distance" library and Pol II "flux" library that will be ligated to vector backbone and performed a test DNA-seq on them. DNA-seq results suggested that the new strategy worked as expected without extensive template switching.



Appendix Figure 5 Design of Pol II "flux" library construction using In-Fusion cloning strategy

The plasmid templates of Pol II "flux" library contain three critical components: UAS, TATA-TSS region, and barcode. These regions can be individually generated as DNA fragment pools, with 20 bp overlaps with adjacent components at their ends. These DNA fragments can be fused and cloned into vector in a single reaction by In-Fusion enzyme via recognizing those overlaps.

Second, the observed poor quality of the first 22 cycles of TSS-seq Read 2 has been determined to be caused by lack of sequence diversity at 3′-end and can be solved by a "stuffing" strategy (see below). Because all RNA products share the same annealing region for reverse transcription of TSS-seq sample preparation, 22 bp sequences at 3′-end of sequencing insert are exactly the same. This lack of diversity where sequencing Read 2 runs risks failing due to loss of signal for cluster identification in Illumina sequencing. To overcome this issue, a mix of Reverse Transcription (RT) primers with different lengths of "stuffing" sequences can be used to add artificial diversity to 3′-end of products for sequencing and therefore to Read 2 (**Appendix Figure 6**). Diversity is added because the mixture of template lengths moves identical sequences into different frames. This "stuffing" strategy has been used for DNA-seq of my TSS sequence specificity libraries and successfully generated high quality reads. Therefore, it would be promising to solve the poor-quality issue of TSS-seq Read 2.



Appendix Figure 6 The "stuffing" strategy to solve no diversity issue

To overcome no sequence diversity at 3′-end of sequencing insert, "stuffing" sequences can be added to 3'-end of final products to artificially create a high diversity. During the 1st strand cDNA synthesis step, a mix of RT primers with different lengths of "stuffing" sequences will be used, so that artificial diversity would be added to 3'-end of final products and therefore to Read 2.

Appendix B Yeast strains, plasmids, and oligonucleotides used in this dissertation

Appendix B.1 Yeast strains

Yeast strains used in **Chapter 2** for Pol II MASTER libraries and **Appendix A** for architecture libraries are listed in **Appendix Table 3**.

Appendix Table 3 Yeast strains used for libraries in Chapter 2 and Appendix A

Strain #	Plasmids	Genotype
CKY954	pCK859 pRS315H3alt-RPB1* XmaI 1122-1123 T69 corrected	*his3Δ200 leu2Δ0 or 1 ura3-52 lys2-128∂ trp1Δ63 rpb1Δ::CLONATMX*
CKY956	pCK887 pRS315H3alt-RPB1* XmaI 1122-1123 H1085Q T69 corrected	*his3Δ200 leu2Δ0 or 1 ura3-52 lys2-128∂ trp1Δ63 rpb1Δ::CLONATMX*
CKY960	pCK871 pRS315H3alt-rpb1* 10-88 (F1086S) T69 corrected	*his3Δ200 leu2Δ0 or 1 ura3-52 lys2-128∂ trp1Δ63 rpb1Δ::CLONATMX*
CKY962	pCK867 pRS315H3alt-RPB1* rpb1 10-110 (G1097D) T69 corrected	*his3Δ200 leu2Δ0 or 1 ura3-52 lys2-128∂ trp1Δ63 rpb1Δ::CLONATMX*
CKY964	pCK960 pRS315H3alt-RPB1* E1103G T69 corrected	*his3Δ200 leu2Δ0 or 1 ura3-52 lys2-128∂ trp1Δ63 rpb1Δ::CLONATMX*

Yeast strains used in **Chapter 3** for *tfb3* mutant screen, *tfb3/ssl2* double mutants, *tfb3/sub1*Δ double mutants, and related experiments are listed in **Appendix Table 4-6**.

Appendix Table 4 Yeast strains used for study of *tfb3* mutant screen in Chapter 3

Strain #	Plasmids	Genotype
CKY2191	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
CKY2196	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY1 / CKY2212	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY2 / CKY2213	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY3 / CKY2214	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*

Appendix Table 4 (continued).

YZY4 / CKY2215	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY16	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY17	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY18	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY19	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY20	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY21	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY22	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY23	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY24	pYZ84 = pRS315 CEN LEU2 - tfb3-84 C13S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY25	pYZ84 = pRS315 CEN LEU2 - tfb3-84 C13S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY28	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY29	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY30	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY31	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY32	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY33	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY34	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY35	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY36	pYZ84 = pRS315 CEN LEU2 - tfb3-84 C13S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY37	pYZ84 = pRS315 CEN LEU2 - tfb3-84 C13S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY40	pYZ7 = pRS315 CEN LEU2 - tfb3-7 V88E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY41	pYZ7 = pRS315 CEN LEU2 - tfb3-7 V88E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY46	pYZ34 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY47	pYZ34 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY50	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*

Appendix Table 4 (continued).

YZY51	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY52	pYZ47 = pRS315 CEN LEU2 - tfb3-47 E112K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY53	pYZ47 = pRS315 CEN LEU2 - tfb3-47 E112K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY58	pYZ86 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY59	pYZ86 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY64	pYZ7 = pRS315 CEN LEU2 - tfb3-7 V88E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY65	pYZ7 = pRS315 CEN LEU2 - tfb3-7 V88E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY70	pYZ34 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY71	pYZ34 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY74	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY75	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY76	pYZ47 = pRS315 CEN LEU2 - tfb3-47 E112K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY77	pYZ47 = pRS315 CEN LEU2 - tfb3-47 E112K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY82	pYZ86 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY83	pYZ86 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY86	pYZ349 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY87	pYZ349 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY92	pYZ392 = pRS315-tfb3-99 K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY93	pYZ392 = pRS315-tfb3-99 K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY94	pYZ392 = pRS315-tfb3-99 K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY95	pYZ392 = pRS315-tfb3-99 K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY126	pYZ1 = pRS315 CEN LEU2 - tfb3-1 E112V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY127	pYZ1 = pRS315 CEN LEU2 - tfb3-1 E112V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY128	pYZ1 = pRS315 CEN LEU2 - tfb3-1 E112V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY129	pYZ1 = pRS315 CEN LEU2 - tfb3-1 E112V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY130	pYZ10 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*

Appendix Table 4 (continued).

YZY131	pYZ10 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY132	pYZ10 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY133	pYZ10 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY134	pYZ19 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY135	pYZ19 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY136	pYZ19 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY137	pYZ19 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY142	pYZ337 = pRS315 CEN LEU2 - tfb3-130 Y56F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY143	pYZ337 = pRS315 CEN LEU2 - tfb3-130 Y56F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY144	pYZ337 = pRS315 CEN LEU2 - tfb3-130 Y56F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY145	pYZ337 = pRS315 CEN LEU2 - tfb3-130 Y56F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY146	pYZ361 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY147	pYZ361 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY148	pYZ361 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY149	pYZ361 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY160	pYZ82 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY161	pYZ82 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY162	pYZ55 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY163	pYZ55 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY164	pYZ60 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY165	pYZ60 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY166	pYZ32 = pRS315 CEN LEU2 - tfb3-32 N66D	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY167	pYZ32 = pRS315 CEN LEU2 - tfb3-32 N66D	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY168	pYZ122 = pRS315 CEN LEU2 - tfb3-106 F68L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY169	pYZ122 = pRS315 CEN LEU2 - tfb3-106 F68L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY170	pYZ127 = pRS315 CEN LEU2 - tfb3-111 F68S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*

Appendix Table 4 (continued).

YZY171	pYZ127 = pRS315 CEN LEU2 - tfb3-111 F68S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY172	pYZ88 = pRS315 CEN LEU2 - tfb3-88 F68V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY173	pYZ88 = pRS315 CEN LEU2 - tfb3-88 F68V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY174	pYZ318 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY175	pYZ318 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY176	pYZ319 = pRS315 CEN LEU2 - tfb3-119 M12K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY177	pYZ319 = pRS315 CEN LEU2 - tfb3-119 M12K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY178	pYZ338 = pRS315 CEN LEU2 - tfb3-131 I15N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY179	pYZ338 = pRS315 CEN LEU2 - tfb3-131 I15N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY180	pYZ335 = pRS315 CEN LEU2 - tfb3-128 I38F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY181	pYZ335 = pRS315 CEN LEU2 - tfb3-128 I38F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY182	pYZ358 = pRS315 CEN LEU2 - tfb3-151 D44G	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY183	pYZ358 = pRS315 CEN LEU2 - tfb3-151 D44G	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY184	pYZ72 = pRS315 CEN LEU2 - tfb3-72 I46T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY185	pYZ72 = pRS315 CEN LEU2 - tfb3-72 I46T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY186	pYZ43 = pRS315 CEN LEU2 - tfb3-43 I62A	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY187	pYZ43 = pRS315 CEN LEU2 - tfb3-43 I62A	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY198	pYZ318 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY199	pYZ318 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY200	pYZ319 = pRS315 CEN LEU2 - tfb3-119 M12K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY201	pYZ319 = pRS315 CEN LEU2 - tfb3-119 M12K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY202	pYZ338 = pRS315 CEN LEU2 - tfb3-131 I15N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY203	pYZ338 = pRS315 CEN LEU2 - tfb3-131 I15N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY204	pYZ335 = pRS315 CEN LEU2 - tfb3-128 I38F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY205	pYZ335 = pRS315 CEN LEU2 - tfb3-128 I38F	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY206	pYZ358 = pRS315 CEN LEU2 - tfb3-151 D44G	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*

Appendix Table 4 (continued).

YZY207	pYZ358 = pRS315 CEN LEU2 - tfb3-151 D44G	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY208	pYZ72 = pRS315 CEN LEU2 - tfb3-72 I46T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY209	pYZ72 = pRS315 CEN LEU2 - tfb3-72 I46T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY210	pYZ43 = pRS315 CEN LEU2 - tfb3-43 I62A	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY211	pYZ43 = pRS315 CEN LEU2 - tfb3-43 I62A	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY240	pYZ82 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY241	pYZ82 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY242	pYZ55 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY243	pYZ55 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY244	pYZ60 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY245	pYZ60 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY246	pYZ32 = pRS315 CEN LEU2 - tfb3-32 N66D	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY247	pYZ32 = pRS315 CEN LEU2 - tfb3-32 N66D	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY248	pYZ122 = pRS315 CEN LEU2 - tfb3-106 F68L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY249	pYZ122 = pRS315 CEN LEU2 - tfb3-106 F68L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY250	pYZ127 = pRS315 CEN LEU2 - tfb3-111 F68S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY251	pYZ127 = pRS315 CEN LEU2 - tfb3-111 F68S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
YZY252	pYZ88 = pRS315 CEN LEU2 - tfb3-88 F68V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY253	pYZ88 = pRS315 CEN LEU2 - tfb3-88 F68V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*

Appendix Table 5 Yeast strains used for study of *tfb3/ssl2* double mutants in Chapter 3

Strain #	**Plasmids**	**Genotype**
YZY1 / CKY2212	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY2 / CKY2213	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3*
YZY3 / CKY2214	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1*
CKY3417	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*

Appendix Table 5 (continued).

CKY3418	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3419	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3420	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3421	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY3422	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY3423	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY3424	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY3610	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3611	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3612	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3613	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3614	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY3615	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY3616	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY3617	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY3618	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY3619	pCK1632= pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY3629	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3630	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3631	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3632	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3633	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**

Appendix Table 5 (continued).

CKY3634	pCK1664 = pRS315 TFB3 CEN LEU2	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY3681	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3682	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3683	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3684	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3685	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY3686	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY3657	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3658	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3659	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3660	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3661	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3662	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3663	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3664	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3665	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY3666	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY3667	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY3668	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY3669	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY3670	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY3671	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*

Appendix Table 5 (continued).

CKY3672	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY3809	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3810	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3811	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3812	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3813	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3814	pCK2534 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY3815	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY3816	pCK2527 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY3934	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3935	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3936	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3937	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3938	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3939	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3940	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3941	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3942	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3943	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3944	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3945	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3946	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3947	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3948	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3949	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3950	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*

Appendix Table 5 (continued).

CKY3951	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3952	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3953	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3954	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3955	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3956	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3957	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3958	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3959	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3960	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3961	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3962	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3963	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3964	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3965	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3966	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3967	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3968	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3969	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3970	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3971	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3972	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3973	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230I*
CKY3974	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3975	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3976	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*

Appendix Table 5 (continued).

CKY3977	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY3978	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3979	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3980	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3981	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3982	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3983	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3984	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3985	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3986	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3987	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3988	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3989	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3990	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3991	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3992	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3993	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3994	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*

Appendix Table 5 (continued).

CKY3995	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3996	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3997	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3998	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY3999	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY4000	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY4001	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY4002	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY4003	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY4004	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY4005	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I*
CKY4006	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4007	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4008	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4009	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4010	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4011	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4012	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4013	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4014	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4015	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*

Appendix Table 5 (continued).

CKY4016	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4017	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4018	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4019	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4020	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4021	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4022	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4023	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4024	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4025	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4026	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4027	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4028	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4029	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4030	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4031	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4032	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4033	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 R636C*
CKY4034	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4035	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4036	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4037	pCK2526 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4038	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*

Appendix Table 5 (continued).

CKY4039	pCK2530 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4040	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4041	pCK2531 = pRS315 CEN LEU2 - tfb3-55 L22P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4042	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4043	pCK2532 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4044	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4045	pCK2533 = pRS315 CEN LEU2 - tfb3-60 S23P	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4046	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4047	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4048	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4049	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4050	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4051	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4052	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4053	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4054	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4055	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4056	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*

Appendix Table 5 (continued).

CKY4057	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4058	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4059	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4060	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4061	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 R636C*
CKY4062	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4063	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4064	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4065	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4066	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4067	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4068	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4069	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4070	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4071	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4072	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4073	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4074	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4075	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4076	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4077	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 V473D*
CKY4078	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4079	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4080	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*

Appendix Table 5 (continued).

CKY4081	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4082	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4083	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4084	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4085	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4086	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4087	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4088	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4089	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4090	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4091	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4092	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4093	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 D522V*
CKY4094	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4095	pCK2525 = pRS315 CEN LEU2 - tfb3-34 L218S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4096	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4097	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4098	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4099	pCK2522 = pRS315 CEN LEU2 - tfb3-10 E131V	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4100	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4101	pCK2523 = pRS315 CEN LEU2 - tfb3-19 Y56C	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4102	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4103	pCK2542 = pRS315 CEN LEU2 - tfb3-118 E7K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4104	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4105	pCK2547 = pRS315 CEN LEU2 - tfb3-142 I62T	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4106	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**

Appendix Table 5 (continued).

CKY4107	pCK2549 = pRS315 CEN LEU2 - tfb3-155 D204N	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4108	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4109	pCK2550 = pRS315 CEN LEU2 - tfb3(-99) K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 Y750**
CKY4110	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY4111	pCK2536 = pRS315 CEN LEU2 - tfb3-82 Y21H	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY4112	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*
CKY4113	pCK2538 = pRS315 CEN LEU2 - tfb3-86 R64K	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1 ssl2 N230D*

Appendix Table 6 Yeast strains used for study of *tfb3/sub1*Δ double mutant in Chapter 3

Strain #	**Plasmids**	**Genotype**
CKY3497	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3498	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3499	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3500	pCK1632 = pRSII316 TFB3 CEN URA3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3505	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3506	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3507	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3508	pYZ59 = pRS315 CEN LEU2 - tfb3-59 C16R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3509	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3510	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3511	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3512	pYZ62 = pRS315 CEN LEU2 - tfb3-62 K17E	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3513	pYZ392 = pRS315 CEN LEU2 - tfb3-99_K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3514	pYZ392 = pRS315 CEN LEU2 - tfb3-99_K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3515	pYZ392 = pRS315 CEN LEU2 - tfb3-99_K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3516	pYZ392 = pRS315 CEN LEU2 - tfb3-99_K17R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*

Appendix Table 6 (continued).

CKY3517	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3518	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3519	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3520	pYZ42 = pRS315 CEN LEU2 - tfb3-42 F28L	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3521	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3522	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3523	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3524	pYZ53 = pRS315 CEN LEU2 - tfb3-53 C39R	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3525	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3526	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3527	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3528	pYZ37 = pRS315 CEN LEU2 - tfb3-37 F73S	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3501	pCK1664 = pRS315 TFB3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3502	pCK1664 = pRS315 TFB3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 sub1Δ::natMX*
CKY3503	pCK1664 = pRS315 TFB3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*
CKY3504	pCK1664 = pRS315 TFB3	*ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hph gal10Δ56 RPB3::TAP::KlacTRP1 sub1Δ::natMX*

Appendix B.2 Plasmids

Plasmids used in **Chapter 2** for Pol II MASTER libraries and **Appendix A** for architecture libraries are listed in **Appendix Table 7**.

Appendix Table 7 Plasmids used for Pol II MASTER and architecture libraries in Chapter 2 and Appendix A

Strain #	Plasmid #	Plasmid	Plasmid Genotype
CKB1316	pCK1316	pRSII413	HIS3 CEN6 ARSH4 ampr f(1) ori, pBluescript II SK+ Backbone
CKB456	pCK456	pYM-N7 ADH Promoter, -natNT2	ampr ColE1 ori
CKB1412	pCK1412	pRS413 VRG4 promoter (TL1 mutation) yeGFP CYC1 terminator	HIS3 CEN ARS ampr ColE1 ori VRG4 promoter (TL1 mutation) yeGFP CYC1 terminator

Plasmids used in **Chapter 3** for *tfb3* mutant screen, *tfb3/ssl2* double mutants, *tfb3/sub1*Δ double mutants, and related experiments are listed in **Appendix Table 8-10**.

Appendix Table 8 Plasmids used for study of *tfb3* mutant screen in Chapter 3

Strain #	Plasmid #	Plasmid	Plasmid Genotype
CKB586	pRS315	pRS315	
CKB1632	pCK1632	pRSII316 TFB3 CEN URA3	URA3 CEN6 ARSH4 ampr f(1) ori TFB3 pBluescript KS+ Backbone
CKB1664	pCK1664	pRS315 TFB3	LEU2 CEN6 ARSH4 ampr f(1) ori, TFB3 pBluescript KS+ Backbone
CKB2520	pCK2520	pRS315_tfb3-1 E112V	ampr ColE1 CEN ARS LEU2
CKB2521	pCK2521	pRS315_tfb3-7 V88E	ampr ColE1 CEN ARS LEU2
CKB2522	pCK2522	pRS315_tfb3-10 E131V	ampr ColE1 CEN ARS LEU2
CKB2523	pCK2523	pRS315_tfb3-19 Y56C	ampr ColE1 CEN ARS LEU2
CKB2524	pCK2524	pRS315_tfb3-32 N66D	ampr ColE1 CEN ARS LEU2
CKB2525	pCK2525	pRS315_tfb3-34 L218S	ampr ColE1 CEN ARS LEU2
CKB2526	pCK2526	pRS315_tfb3-37 F73S	ampr ColE1 CEN ARS LEU2
CKB2527	pCK2527	pRS315_tfb3-42 F28L	ampr ColE1 CEN ARS LEU2
CKB2528	pCK2528	pRS315_tfb3-43 I62A	ampr ColE1 CEN ARS LEU2
CKB2529	pCK2529	pRS315_tfb3-47 E112K	ampr ColE1 CEN ARS LEU2
CKB2530	pCK2530	pRS315_tfb3-53 C39R	ampr ColE1 CEN ARS LEU2
CKB2531	pCK2531	pRS315_tfb3-55 L22P	ampr ColE1 CEN ARS LEU2
CKB2532	pCK2532	pRS315_tfb3-59 C16R	ampr ColE1 CEN ARS LEU2
CKB2533	pCK2533	pRS315_tfb3-60 S23P	ampr ColE1 CEN ARS LEU2
CKB2534	pCK2534	pRS315_tfb3-62 K17E	ampr ColE1 CEN ARS LEU2
CKB2535	pCK2535	pRS315_tfb3-72 I46T	ampr ColE1 CEN ARS LEU2
CKB2536	pCK2536	pRS315_tfb3-82 Y21H	ampr ColE1 CEN ARS LEU2

Appendix Table 8 (continued).

CKB2537	pCK2537	pRS315_tfb3-84 C13S	ampr ColE1 CEN ARS LEU2
CKB2538	pCK2538	pRS315_tfb3-86 R64K	ampr ColE1 CEN ARS LEU2
CKB2539	pCK2539	pRS315_tfb3-88 F68V	ampr ColE1 CEN ARS LEU2
CKB2550	pCK2550	pRS315-tfb3(-99) K17R	ampr ColE1 CEN ARS LEU2
CKB2540	pCK2540	pRS315_tfb3-106 F68L	ampr ColE1 CEN ARS LEU2
CKB2541	pCK2541	pRS315_tfb3-111 F68S	ampr ColE1 CEN ARS LEU2
CKB2542	pCK2542	pRS315_tfb3-118 E7K	ampr ColE1 CEN ARS LEU2
CKB2543	pCK2543	pRS315_tfb3-119 M12K	ampr ColE1 CEN ARS LEU2
CKB2544	pCK2544	pRS315_tfb3-128 I38F	ampr ColE1 CEN ARS LEU2
CKB2545	pCK2545	pRS315_tfb3-130 Y56F	ampr ColE1 CEN ARS LEU2
CKB2546	pCK2546	pRS315_tfb3-131 I15N	ampr ColE1 CEN ARS LEU2
CKB2547	pCK2547	pRS315_tfb3-142 I62T	ampr ColE1 CEN ARS LEU2
CKB2548	pCK2548	pRS315_tfb3-151 D44G	ampr ColE1 CEN ARS LEU2
CKB2549	pCK2549	pRS315_tfb3-155 D204N	ampr ColE1 CEN ARS LEU2

Appendix Table 9 Plasmids used for study of *tfb3/ssl2* double mutants in Chapter 3

Strain #	**Plasmid #**	**Plasmid**	**Plasmid Genotype**
CKB2070	pCK2070	pRS306 ssl2 T176P	URA3 ampr ColE1 ori, ssl2 T176P
CKB2071	pCK2071	pRS306 ssl2 L225P	URA3 ampr ColE1 ori, ssl2 L225P
CKB2072	pCK2072	pRS306 ssl2 N230I	URA3 ampr ColE1 ori, ssl2 N230I
CKB2073	pCK2073	pRS306 ssl2 N230D	URA3 ampr ColE1 ori, ssl2 N230D
CKB2074	pCK2074	pRS306 ssl2 V473D	URA3 ampr ColE1 ori, ssl2 V473D
CKB2075	pCK2075	pRS306 ssl2 D522V	URA3 ampr ColE1 ori, ssl2 D522V
CKB2076	pCK2076	pRS306 ssl2 R636C	URA3 ampr ColE1 ori, ssl2 R636C

Appendix Table 10 Plasmids used for study of *tfb3/sub1*Δ double mutants in Chapter 3

Strain #	**Plasmid #**	**Plasmid**	**Plasmid Genotype**
CKB236	pCK236	pBS KS+RS natMX	pBS KS backbone ampr ColE1 ori

Appendix B.3 Oligonucleotides

Oligonucleotides used in **Chapter 2** for Pol II MASTER libraries are listed in **Appendix Table 11**.

Appendix Table 11 Oligonucleotides used for Pol II MASTER libraries in Chapter 2

Oligo #	Name	Sequence	Function
CKO2447	SNR37	CTGATGTTAGCCTTTTGAAAAT GTYRNNNNNNTGAAAAAAAAT CAAAAGAAAAATTTG	SNR37-TSS AYR lib R
CKO2448	GFP-CYC1 fusion (pCK1412 etc)	AAAGGCTAACATCAGNNNNAN NNNCNNNNTNNNNGNNNNATG TCTAAAGGTGAAGAATT	GFP-CYCterm-BAR-SacI F
CKO2449	SNR37	CTGATGTTAGCCTTTTGAAAAT GTRYNNNNNNTGAAAAAAAAT CAAAAGAAAAATTTG	SNR37-TSS ARY lib R
CKO2450	SNR37	CTGATGTTAGCCTTTTGAAAAT GTYRNNNNNNVGAAAAAAAAT CAAAAGAAAAATTTG	SNR37-TSS BYR lib R
CKO1608	GAL1 UAS MASTER product ΔSacI	GGTACCGGGCCCCCCCTCGAGT CCTAGTACGGATTAGAAG	CKO1577 sewing F ΔSacI
CKO1578	GAL1/pYM-N22	GGACAACACTAGCTCTATTTAT ATCTGTTAATAGATCAAAAATC	GAL1 UAS-XhoI R
CKO1579	SNR37	GATTTTTGATCTATTAACAGAT ATAAATAGAGCTAGTGTTGTCC	SNR37-TSS lib F
CKO1582	GFP-CYC1 fusion (pCK1412 etc)	GGGAACAAAAGCTGGAGCTCC GAGCGTCCCAAAACCTTCTC	GFP-CYCterm-BAR-SacI R
CKO1586	SNR37 MASTER product	CTGATGTTAGCCTTTTGAAAAT GT	CKO1580 sewing R
CKO1588	CYC1t MASTER product	GGGAACAAAAGCTGGAGCTC	CKO1582 sewing R
CKO2539	stuf_1-SNR37-F	GTGACGACTGTAGCTCAGTTCA AATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end
CKO2540	stuf_2-SNR37-F	AATCTCTAGCCGTAAGTCATTC AAATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end
CKO2541	stuf_3-SNR37-F	CGCTGACGATATCGCAGTGCTT CAAATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end
CKO2542	stuf_4-SNR37-F	TCAGATGTCAGCATGTCATAGT TCAAATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end
CKO2543	stuf_5-SNR37-F	CTATCGCTAGATCGCATGCGAG TTCAAATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end

Appendix Table 11 (continued).

CKO2544	stuf_6-SNR37-F	GCGCGTACGTCAGTTCGACAGAGTTCAAATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end
CKO2545	stuf_7-SNR37-F	AGTAACTGCATGACAGCTAGCGCGTTCAAATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end
CKO2546	stuf_8-SNR37-F	TACGTAGATCGCTAGTACGCACGCGTTCAAATTTTTCTTTTGATTTT	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 5′-end
CKO2547	TruSeq_P5_end_adpter-stuf_1-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGACGACTGTAGCTCAG	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2548	TruSeq_P5_end_adpter-stuf_2-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAATCTCTAGCCGTAAGTCA	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2549	TruSeq_P5_end_adpter-stuf_3-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTGACGATATCGCAGTGC	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2550	TruSeq_P5_end_adpter-stuf_4-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAGATGTCAGCATGTCATA	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2551	TruSeq_P5_end_adpter-stuf_5-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTATCGCTAGATCGCATGCG	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2552	TruSeq_P5_end_adpter-stuf_6-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGCGTACGTCAGTTCGACA	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2553	TruSeq_P5_end_adpter-stuf_7-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTAACTGCATGACAGCTAG	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2554	TruSeq_P5_end_adpter-stuf_8-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACGTAGATCGCTAGTACGC	Add TruSeq P5 end adpater sequence to stuffered amplicom 5′-end
CKO2555	stuf_1-GFP-R	GTGACTCAGTGAATAATTCTTCACCTTTA	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 3′-end
CKO2556	stuf_2-GFP-R	TGACTCAGTGAATAATTCTTCACCTTTA	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 3′-end
CKO2557	stuf_3-GFP-R	GACTCAGTGAATAATTCTTCACCTTTA	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 3′-end
CKO2558	stuf_4-GFP-R	ACTCAGTGAATAATTCTTCACCTTTA	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 3′-end
CKO2559	stuf_5-GFP-R	CTCAGTGAATAATTCTTCACCTTTA	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 3′-end
CKO2560	stuf_6-GFP-R	TCAGTGAATAATTCTTCACCTTTA	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 3′-end

Appendix Table 11 (continued).

CKO2561	stuf_7-GFP-R	CAGTGAATAATTCTTCACCTTTA	Add stuffer sequences to MASTER Pol II sequence specificity libraries amplicon 3′-end
CKO2126	MASTER library R	AGTGAATAATTCTTCACCTTTA	Amplify library for sequencing
CKO2562	TruSeq_P7_end_adpter-stuf_1-GFP-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGACTCAGTGAATAATTCT	Add TruSeq P7 end adpater sequence to stuffered amplicom 3′-end
CKO2563	TruSeq_P7_end_adpter-stuf_2-GFP-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGACTCAGTGAATAATTCTT	Add TruSeq P7 end adpater sequence to stuffered amplicom 3′-end
CKO2564	TruSeq_P7_end_adpter-stuf_3-GFP-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGACTCAGTGAATAATTCTTC	Add TruSeq P7 end adpater sequence to stuffered amplicom 3′-end
CKO2565	TruSeq_P7_end_adpter-stuf_4-GFP-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACTCAGTGAATAATTCTTCA	Add TruSeq P7 end adpater sequence to stuffered amplicom 3′-end
CKO2566	TruSeq_P7_end_adpter-stuf_5-GFP-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTCAGTGAATAATTCTTCAC	Add TruSeq P7 end adpater sequence to stuffered amplicom 3′-end
CKO2567	TruSeq_P7_end_adpter-stuf_6-GFP-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCAGTGAATAATTCTTCACC	Add TruSeq P7 end adpater sequence to stuffered amplicom 3′-end
CKO2568	TruSeq_P7_end_adpter-stuf_7-GFP-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGTGAATAATTCTTCACCT	Add TruSeq P7 end adpater sequence to stuffered amplicom 3′-end
CKO2284	P7_stuf_2126 R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGTGAATAATTCTTCACCTTTA	Add sequencing adapter and stuffer to MASTER v3 library amplicon

Oligonucleotides used in **Chapter 3** for *tfb3* mutant screen, *tfb3/ssl2* double mutants, *tfb3/sub1*Δ double mutants, and related experiments are listed in **Appendix Table 12-14**.

Appendix Table 12 Oligonucleotides used for study of *tfb3* mutant screen in Chapter 3

Oligo #	**Name**	**Sequence**	**Function**
CKO1694	TFB3-XhoI F	CCCAAACTCGAGTGTCGACTTGCTCGTAACAT	Cloning TFB3
CKO1695	TFB3-SacI R	CCCAAAGAGCTCCTGGGTTTCATTTAACATGTA	Cloning TFB3
CKO1696	tfb3Δ F	AAATTTGAGTTTGTTTCCTCTTAGGATTTGCAGCATTAGTGCGTACGCTGCAGGTCGAC	tfb3 Toolkit KO
CKO1697	tfb3Δ R	GATTACATAGCTTATGCCACGTTGCACACTATCCTGATGTTATCGATGAATTCGAGCTCG	tfb3 Toolkit KO

Appendix Table 12 (continued).

CKO1701	TFB3 3′ R	CTGAAACAGATCGTGAATGC	Verification of C-term tags/KO/sequencing
CKO1702	TFB3 5′ F	CAGCTAAGAAACTGAACTGCTT	Verification of KO/sequencing
CKO1799	TFB3 Gap-repair R	GAACACTGAAACAGATCGTG	Tfb3 gapping PCR
CKO1800	TFB3 Gap-repair F	TGTCGACTTGCTCGTAACAT	Tfb3 gapping PCR
CKO1801	TFB3 Gap Seq1 R	AAGGTATGGGCCAACCGTAG	Tfb3 gapping sequencing
CKO1802	TFB3 Gap Seq2 F	ATGCTTATGGATGAGTATG	Tfb3 gapping sequencing
CKO1803	TFB3 Gap Seq3 F	AGAGATAGAAAGTTTCGAACA	Tfb3 gapping sequencing
CKO2127	TFB3 K17R F	GTCCGATCTGTAGGACAGATCG ATACCTTTC	Site-directed mutagenesis
CKO2128	TFB3 K17R R	CGATCTGTCCTACAGATCGGAC ACATGTCC	Site-directed mutagenesis
CKO401	ADH1 PE	ATAACACCTTTTTGAGTTTCTG G	Primer extension

Appendix Table 13 Oligonucleotides used for study of *tfb3/ssl2* double mutants in Chapter 3

Oligo #	**Name**	**Sequence**	**Function**
CKO1184	SSL2 R	CAGCGACGAAAGAATAAGAAG ACGG	KO verification/sequencing
CKO1186	SSL2.2 +1100 ORF F	TCGCTGAGTAAAATGTTTGGTA ATGG	sequencing
CKO1187	SSL2.3 +1700 ORF F	GTATTTAAGAGAAACTGCAAG GAAAAG	sequencing
CKO1293	SSL2 Gap-repair Mid R	GCATATTCTTGTAAGGCGTA	
CKO1510	SSL2_Gap1_Seq_ R	AAGGAGTGAACGGCATCAAT	SSL2 mutant sequencing, mutants screened from gapping test
CKO1511	SSL2_Gap2_Seq_ F	AGACCAGACCATGCTTCCAG	SSL2 mutant sequencing, mutants screened from gapping test

Appendix Table 14 Oligonucleotides used for study of *tfb3/sub1*Δ double mutants in Chapter 3

Oligo #	**Name**	**Sequence**	**Function**
CKO512	sub1Δ F	GTCCCCAATCAAAGAAACACTTTTGGTTGC CCACTTAAGAGATTGTACTGAGAGTGCAC	pRS KO
CKO513	sub1Δ R	CTCTAATAAGCTCGTTGGATGGAAGACGTT GACATAAGCCTGTGCGGTATTTCACACCG	pRS KO
CKO534	SUB1 5′ Promoter F	TTGGCTTGCCTTCTGCTCTC	Verification of N-terminal tagging, 5′ insertions, KO
CKO473	SUB1 R	GTACTAGTTGTACGGGGAAAATG	Verification of KO/Amplification of gene

Oligonucleotides used in **Appendix A** for architecture libraries are listed in **Appendix Table 15**.

Appendix Table 15 Oligonucleotides used for archtecture libraries in Appendix A

Oligo #	Name	Sequence	Function
CKO1582	GFP-CYC1 fusion (pCK1412 etc)	GGGAACAAAAGCTGGAGCTCCG AGCGTCCCAAAACCTTCTC	GFP-CYCterm-BAR-SacI R
CKO1588	CYC1t MASTER product	GGGAACAAAAGCTGGAGCTC	CKO1582 sewing R
CKO2591	ADH1 UAS overlapping F	GCTATCAAGTATAAATAGAC	ADH1 -137 to -118; Add flanking region of ADH1 to ADH1 TATA-TSS distance library
CKO2592	ADH1-Barcode overlapping R	TGTATATGAGATAGTTGATTGTA TGCTTGGTATAGCTTGAAATATT GT	ADH1 -48 to -1; Add flanking region of ADH1 to ADH1 TATA-TSS distance library
CKO2593	ADH1 UAS-XhoI F	GGTACCGGGCCCCCCCTCGAGGC ATGCAACTTCTTTTCTTT	ADH1 UAS-XhoI F
CKO2596	CKO2593 CKO2599 sewing F	GGTACCGGGCCCCCCCTCGAG	CKO2593 CKO2599 sewing F
CKO2597	RPS5 UAS overlapping F	GTAGGTAATGGATTATCCACAAT GTGCCCATTAGTTCTAAAATATTT TGTACTTTTTAT	RPS5 -162 to -104; Add flanking region of RPS5 to RPS5 distance library
CKO2598	RPS5-Barcode overlapping R	CTTTGGAATGGTCGGTTATTTCTA GTCTCTTTTCTTGGTATTATGA	RPS5 -46 to -1; Add flanking region of RPS5 to RPS5 distance library
CKO2599	RPS5 UAS-XhoI F	GGTACCGGGCCCCCCCTCGAGTC AAGACCTTTTATACAACA	RPS5 UAS-XhoI F
CKO2602	ADH1-Barcode_v2-GFP F	AATCAACTATCTCATATACANNN NNANNNNNCNNNNNATGTCTAA AGGTGAAGAATTAT	ADH1-Barcode_v2-GFP F
CKO2603	RPS5-Barcode_v2-GFP F	AATAACCGACCATTCCAAAGNNN NNTNNNNNGNNNNNATGTCTAA AGGTGAAGAATTAT	RPS5-Barcode_v2-GFP F
CKO2604	WT_FD-Barcode_v2-GFP F	TTTCAAAAGGCTAACATCAGNNN NNCNNNNNTNNNNNATGTCTAA AGGTGAAGAATTAT	WT_FD-Barcode_v2-GFP F
CKO2605	Mutated_FD-Barcode_v2-GFP F	TTTCAAAAGGCTTTCTTCTGNNN NNANNNNNGNNNNNATGTCTAA AGGTGAAGAATTAT	Mutated_FD-Barcode_v2-GFP F
CKO2606	AHD1-handle F	TTTGCTGTCTTGCTATCAAG	Inner PCR handles for Pol II Flux Detector library; ADH1 UAS overlapping -148 to -129
CKO2607	CYC1-handle F	GAGCAGATCCGCCAGGCGTG	Inner PCR handles for Pol II Flux Detector library; CYC1 UAS overlapping -247 to -228

Appendix Table 15 (continued).

CKO2608	ERG8-handle F	TGTCGAATGGAAAGAAAGAT	Inner PCR handles for Pol II Flux Detector library; ERG8 UAS overlapping -100 to -81
CKO2609	RPS5-handle F	TTTGCTGGGAAATCTCCTTT	Inner PCR handles for Pol II Flux Detector library; RPS5 UAS overlapping -265 to -246
CKO2610	HIS4-handle F	ATGAACAGTAGTATACTGTG	Inner PCR handles for Pol II Flux Detector library; HIS4 UAS overlapping -143 to -124
CKO2611	CYC3-handle F	TTATCAAAAAATTTCTCAGC	Inner PCR handles for Pol II Flux Detector library; CYC3 UAS overlapping -125 to -106
CKO2612	WT_FD R	CTGATGTTAGCCTTTTGAAA	Inner PCR handles for Pol II Flux Detector library; WT FD (SNR37)
CKO2613	Mutated_FD R	CAGAAGAAAGCCTTTTGAAA	Inner PCR handles for Pol II Flux Detector library; Mutated FD (SNR37)
CKO2614	WT_FD-YBR135W R	CTGATGTTAGCCTTTTGAAACCAAGAACTTACGACAGAGT	Inner PCR handles R for YBR135W of Pol II FD library; Add WT FD
CKO2615	Mutated_FD-YBR135W R	CAGAAGAAAGCCTTTTGAAACCAAGAACTTACGACAGAGT	Inner PCR handles R for YBR135W of Pol II FD library; Add Mutated FD
CKO2616	WT_FD-YPL135W R	CTGATGTTAGCCTTTTGAAATGTTTTCCTTATTCAATTGT	Inner PCR handles R for YPL135W of Pol II FD library; Add WT FD
CKO2617	Mutated_FD-YPL135W R	CAGAAGAAAGCCTTTTGAAATGTTTTCCTTATTCAATTGT	Inner PCR handles R for YPL135W of Pol II FD library; Add Mutated FD
CKO2618	WT_FD-YNL111C R	CTGATGTTAGCCTTTTGAAATGATTGATTTTCTTTACTAT	Inner PCR handles R for YNL111C of Pol II FD library; Add WT FD
CKO2619	Mutated_FD-YNL111C R	CAGAAGAAAGCCTTTTGAAATGATTGATTTTCTTTACTAT	Inner PCR handles R for YNL111C of Pol II FD library; Add Mutated FD
CKO2620	WT_FD-YJR080C R	CTGATGTTAGCCTTTTGAAATGGATATCCTTTTCTGTGCT	Inner PCR handles R for YJR080C of Pol II FD library; Add WT FD
CKO2621	Mutated_FD-YJR080C R	CAGAAGAAAGCCTTTTGAAATGGATATCCTTTTCTGTGCT	Inner PCR handles R for YJR080C of Pol II FD library; Add Mutated FD
CKO2622	WT_FD-YAL005C R	CTGATGTTAGCCTTTTGAAATGATTACTTTTCTTTTGATG	Inner PCR handles R for YAL005C of Pol II FD library; Add WT FD
CKO2623	Mutated_FD-YAL005C R	CAGAAGAAAGCCTTTTGAAATGATTACTTTTCTTTTGATG	Inner PCR handles R for YAL005C of Pol II FD library; Add Mutated FD
CKO2624	WT_FD-YER124C R	CTGATGTTAGCCTTTTGAAACACTATCACACCTACAATCA	Inner PCR handles R for YER124C of Pol II FD library; Add WT FD
CKO2625	Mutated_FD-YER124C R	CAGAAGAAAGCCTTTTGAAACACTATCACACCTACAATCA	Inner PCR handles R for YER124C of Pol II FD library; Add Mutated FD
CKO2626	WT_FD-YDR508C R	CTGATGTTAGCCTTTTGAAAGATAGTTTTTTTTTCTTACA	Inner PCR handles R for YDR508C of Pol II FD library; Add WT FD
CKO2627	Mutated_FD-YDR508C R	CAGAAGAAAGCCTTTTGAAAGATAGTTTTTTTTTCTTACA	Inner PCR handles R for YDR508C of Pol II FD library; Add Mutated FD
CKO2628	WT_FD-YPR109W R	CTGATGTTAGCCTTTTGAAATGATTTCCTTGCAATAACTA	Inner PCR handles R for YPR109W of Pol II FD library; Add WT FD
CKO2629	Mutated_FD-YPR109W R	CAGAAGAAAGCCTTTTGAAATGATTTCCTTGCAATAACTA	Inner PCR handles R for YPR109W of Pol II FD library; Add Mutated FD
CKO2630	ADH1-YKL106W F	TTTGCTGTCTTGCTATCAAGTATATAAATTGCTAGCAAGT	Inner PCR handles F for YKL106W of Pol II FD library; Add ADH1 UAS overlapping

Appendix Table 15 (continued).

CKO2631	CYC1-YKL106W F	GAGCAGATCCGCCAGGCGTGTATATAAATTGCTAGCAAGT	Inner PCR handles F for YKL106W of Pol II FD library; Add CYC1 UAS overlapping
CKO2632	ERG8-YKL106W F	TGTCGAATGGAAAGAAAGATTATATAAATTGCTAGCAAGT	Inner PCR handles F for YKL106W of Pol II FD library; Add ERG8 UAS overlapping
CKO2633	RPS5-YKL106W F	TTTGCTGGGAAATCTCCTTTTATATAAATTGCTAGCAAGT	Inner PCR handles F for YKL106W of Pol II FD library; Add RPS5 UAS overlapping
CKO2634	HIS4-YKL106W F	ATGAACAGTAGTATACTGTGTATATAAATTGCTAGCAAGT	Inner PCR handles F for YKL106W of Pol II FD library; Add HIS4 UAS overlapping
CKO2635	CYC3-YKL106W F	TTATCAAAAAATTTCTCAGCTATATAAATTGCTAGCAAGT	Inner PCR handles F for YKL106W of Pol II FD library; Add CYC3 UAS overlapping
CKO2636	ADH1-YBL003C F	TTTGCTGTCTTGCTATCAAGTATATAAGGGATGAAGATGT	Inner PCR handles F for YBL003C of Pol II FD library; Add ADH1 UAS overlapping
CKO2637	CYC1-YBL003C F	GAGCAGATCCGCCAGGCGTGTATATAAGGGATGAAGATGT	Inner PCR handles F for YBL003C of Pol II FD library; Add CYC1 UAS overlapping
CKO2638	ERG8-YBL003C F	TGTCGAATGGAAAGAAAGATTATATAAGGGATGAAGATGT	Inner PCR handles F for YBL003C of Pol II FD library; Add ERG8 UAS overlapping
CKO2639	RPS5-YBL003C F	TTTGCTGGGAAATCTCCTTTTATATAAGGGATGAAGATGT	Inner PCR handles F for YBL003C of Pol II FD library; Add RPS5 UAS overlapping
CKO2640	HIS4-YBL003C F	ATGAACAGTAGTATACTGTGTATATAAGGGATGAAGATGT	Inner PCR handles F for YBL003C of Pol II FD library; Add HIS4 UAS overlapping
CKO2641	CYC3-YBL003C F	TTATCAAAAAATTTCTCAGCTATATAAGGGATGAAGATGT	Inner PCR handles F for YBL003C of Pol II FD library; Add CYC3 UAS overlapping
CKO2642	ADH1-YMR195W F	TTTGCTGTCTTGCTATCAAGTATATATGTAGCTAGACACGTTCATAATTTTTATGCTCT	Inner PCR handles F for YMR195W of Pol II FD library; Add ADH1 UAS overlapping
CKO2643	CYC1-YMR195W F	GAGCAGATCCGCCAGGCGTGTATATATGTAGCTAGACACGTTCATAATTTTTATGCTCT	Inner PCR handles F for YMR195W of Pol II FD library; Add CYC1 UAS overlapping
CKO2644	ERG8-YMR195W F	TGTCGAATGGAAAGAAAGATTATATATGTAGCTAGACACGTTCATAATTTTTATGCTCT	Inner PCR handles F for YMR195W of Pol II FD library; Add ERG8 UAS overlapping
CKO2645	RPS5-YMR195W F	TTTGCTGGGAAATCTCCTTTTATATATGTAGCTAGACACGTTCATAATTTTTATGCTCT	Inner PCR handles F for YMR195W of Pol II FD library; Add RPS5 UAS overlapping
CKO2646	HIS4-YMR195W F	ATGAACAGTAGTATACTGTGTATATATGTAGCTAGACACGTTCATAATTTTTATGCTCT	Inner PCR handles F for YMR195W of Pol II FD library; Add HIS4 UAS overlapping
CKO2647	CYC3-YMR195W F	TTATCAAAAAATTTCTCAGCTATATATGTAGCTAGACACGTTCATAATTTTTATGCTCT	Inner PCR handles F for YMR195W of Pol II FD library; Add CYC3 UAS overlapping
CKO2648	ADH1-YOR344C F	TTTGCTGTCTTGCTATCAAGTATATAAACTTCTTTTGTTATCTCCAAAATTTAAACTTA	Inner PCR handles F for YOR344C of Pol II FD library; Add ADH1 UAS overlapping

Appendix Table 15 (continued).

CKO2649	CYC1-YOR344C F	GAGCAGATCCGCCAGGCGTGTATATAAACTTCTTTTGTTATCTCCAAAATTTAAACTTA	Inner PCR handles F for YOR344C of Pol II FD library; Add CYC1 UAS overlapping
CKO2650	ERG8-YOR344C F	TGTCGAATGGAAAGAAAGATTATATAAACTTCTTTTGTTATCTCCAAAATTTAAACTTA	Inner PCR handles F for YOR344C of Pol II FD library; Add ERG8 UAS overlapping
CKO2651	RPS5-YOR344C F	TTTGCTGGGAAATCTCCTTTTATATAAACTTCTTTTGTTATCTCCAAAATTTAAACTTA	Inner PCR handles F for YOR344C of Pol II FD library; Add RPS5 UAS overlapping
CKO2652	HIS4-YOR344C F	ATGAACAGTAGTATACTGTGTATATAAACTTCTTTTGTTATCTCCAAAATTTAAACTTA	Inner PCR handles F for YOR344C of Pol II FD library; Add HIS4 UAS overlapping
CKO2653	CYC3-YOR344C F	TTATCAAAAAATTTCTCAGCTATATAAACTTCTTTTGTTATCTCCAAAATTTAAACTTA	Inner PCR handles F for YOR344C of Pol II FD library; Add CYC3 UAS overlapping
CKO2654	WT_FD-YKL106W R	CTGATGTTAGCCTTTTGAAATAGTAATTTTCAAATGAGAA	Inner PCR handles R for YKL106W of Pol II FD library; Add WT FD
CKO2655	Mutated_FD-YKL106W R	CAGAAGAAAGCCTTTTGAAATAGTAATTTTCAAATGAGAA	Inner PCR handles R for YKL106W of Pol II FD library; Add Mutated FD
CKO2656	WT_FD-YBL003C R	CTGATGTTAGCCTTTTGAAATACTTTCCTAATTCTTATGTAAAAA	Inner PCR handles R for YBL003C of Pol II FD library; Add WT FD
CKO2657	Mutated_FD-YBL003C R	CAGAAGAAAGCCTTTTGAAATACTTTCCTAATTCTTATGTAAAAA	Inner PCR handles R for YBL003C of Pol II FD library; Add Mutated FD
CKO2658	WT_FD-YMR195W R	CTGATGTTAGCCTTTTGAAATGGTTTTATTGTATGTGTTATATGAGAAATTCAATTC	Inner PCR handles R for YMR195W of Pol II FD library; Add WT FD
CKO2659	Mutated_FD-YMR195W R	CAGAAGAAAGCCTTTTGAAATGGTTTTATTGTATGTGTTATATGAGAAATTCAATTC	Inner PCR handles R for YMR195W of Pol II FD library; Add Mutated FD
CKO2660	WT_FD-YOR344C R	CTGATGTTAGCCTTTTGAAATGTGGTGCTTTACGCTTAATAAGAAAGAAAAT	Inner PCR handles R for YOR344C of Pol II FD library; Add WT FD
CKO2661	Mutated_FD-YOR344C R	CAGAAGAAAGCCTTTTGAAATGTGGTGCTTTACGCTTAATAAGAAAGAAAAT	Inner PCR handles R for YOR344C of Pol II FD library; Add Mutated FD
CKO2663	ADH1 UAS R	CTTGATAGCAAGACAGCAAA	ADH1 UAS R
CKO2665	CYC1 UAS R	CACGCCTGGCGGATCTGCTC	CYC1 UAS R
CKO2667	ERG8 UAS R	ATCTTTCTTTCCATTCGACA	ERG8 UAS R
CKO2669	RPS5 UAS R	AAAGGAGATTTCCCAGCAAA	RPS5 UAS R
CKO2671	HIS4 UAS R	CACAGTATACTACTGTTCAT	HIS4 UAS R
CKO2673	CYC3 UAS R	GCTGAGAAATTTTTTGATAA	CYC3 UAS R
CKO3127	ADH1_-118_to_-137 R	GTCTATTTATACTTGATAGC	ADH1 -137 to -118
CKO3128	RPS5_-143_to_-162 R	GTGGATAATCCATTACCTAC	RPS5 -162 to -143
CKO3129	ADH1 UAS-XbaI F	GGTACCGGGCCCCCCTCTAGAGCATGCAACTTCTTTTCTTT	ADH1 UAS-XbaI F
CKO3130	CYC1 UAS-XbaI F	GGTACCGGGCCCCCCTCTAGACACAGAAAAGAATGCAGAAA	CYC1 UAS-XbaI F
CKO3131	ERG8 UAS-XbaI F	GGTACCGGGCCCCCCTCTAGAAGATCTTCCATCTGATCCTC	ERG8 UAS-XbaI F
CKO3132	RPS5 UAS-XbaI F	GGTACCGGGCCCCCCTCTAGATCAAGACCTTTTATACAACA	RPS5 UAS-XbaI F

Appendix Table 15 (continued).

CKO3133	HIS4 UAS-XbaI F	GGTACCGGGCCCCCCTCTAGATTTTTGACAAAACCACAAGA	HIS4 UAS-XbaI F
CKO3134	CYC3 UAS-XbaI F	GGTACCGGGCCCCCCTCTAGAAGGAATATCTAATACTAAGT	CYC3 UAS-XbaI F
CKO3135	CKO3129-3134 sewing F	GGTACCGGGCCCCCCTCTAGA	CKO3129-3134 sewing F

Bibliography

Aanes, H., Ostrup, O., Andersen, I.S., Moen, L.F., Mathavan, S., Collas, P., and Alestrom, P. (2013). Differential transcript isoform usage pre- and post-zygotic genome activation in zebrafish. BMC Genomics *14*, 331.

Abdella, R., Talyzina, A., Chen, S., Inouye, C.J., Tjian, R., and He, Y. (2021). Structure of the human Mediator-bound transcription preinitiation complex. Science *372*, 52-56.

Aibara, S., Schilbach, S., and Cramer, P. (2021). Structures of mammalian RNA polymerase II pre-initiation complexes. Nature *594*, 124-128.

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).

Arkhipova, I.R. (1995). Promoter elements in Drosophila melanogaster revealed by sequence analysis. Genetics *139*, 1359-1369.

Baek, I., Friedman, L.J., Gelles, J., and Buratowski, S. (2021). Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. Mol Cell.

Bangur, C.S., Pardee, T.S., and Ponticelli, A.S. (1997). Mutational analysis of the D1/E1 core helices and the conserved N-terminal region of yeast transcription factor IIB (TFIIB): identification of an N-terminal mutant that stabilizes TATA-binding protein-TFIIB-DNA complexes. Mol Cell Biol *17*, 6784-6793.

Bansal, M., Kumar, A., and Yella, V.R. (2014). Role of DNA sequence based structural features of promoters in transcription initiation and gene expression. Curr Opin Struct Biol *25*, 77-85.

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.

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.

Batut, P., Dobin, A., Plessy, C., Carninci, P., and Gingeras, T.R. (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res *23*, 169-180.

Benanti, J.A., Matyskiela, M.E., Morgan, D.O., and Toczyski, D.P. (2009). Functionally distinct isoforms of Cik1 are differentially regulated by APC/C-mediated proteolysis. Mol Cell *33*, 581-590.

Benatuil, L., Perez, J.M., Belk, J., and Hsieh, C.M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng Des Sel *23*, 155-159.

Bernard, D.G., Gabilly, S.T., Dujardin, G., Merchant, S., and Hamel, P.P. (2003). Overlapping specificities of the mitochondrial cytochrome c and c1 heme lyases. J Biol Chem *278*, 49732-49742.

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. Molecular and Cellular Biology *14*, 226-237.

Blazeck, J., Garg, R., Reed, B., and Alper, H.S. (2012). Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnol Bioeng *109*, 2884-2895.

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.

Boyd, M., Thodberg, M., Vitezic, M., Bornholdt, J., Vitting-Seerup, K., Chen, Y., Coskun, M., Li, Y., Lo, B.Z.S., Klausen, P., *et al.* (2018). Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. Nat Commun *9*, 1661.

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.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol *212*, 563-578.

Bunick, D., Zandomeni, R., Ackerman, S., and Weinmann, R. (1982). Mechanism of RNA polymerase II--specific initiation of transcription in vitro: ATP requirement and uncapped runoff transcripts. Cell *29*, 877-886.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549-561.

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.

Bushnell DA, W.K., Davis RE, Kornberg RD (2005). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science *303*, 983-988.

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*, e28869.

Cairns, B.R., Henry, N.L., and Kornberg, R.D. (1996). TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. Mol Cell Biol *16*, 3308-3316.

Carlson, M., and Botstein, D. (1982). Two differentially regulated mRNAs with different 5' ends encode secreted with intracellular forms of yeast invertase. Cell *28*, 145-154.

Carmel, I., Tal, S., Vig, I., and Ast, G. (2004). Comparative analysis detects dependencies among the 5' splice-site positions. RNA *10*, 828-840.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., *et al.* (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-1563.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., *et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Chatton, B., Walter, P., Ebel, J.P., Lacroute, F., and Fasiolo, F. (1988). The yeast VAS1 gene encodes both mitochondrial and cytoplasmic valyl-tRNA synthetases. J Biol Chem *263*, 52-57.

Chee, M.K., and Haase, S.B. (2012). New and Redesigned pRS Plasmid Shuttle Vectors for Genetic Manipulation of Saccharomycescerevisiae. G3 (Bethesda) *2*, 515-526.

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. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. Mol Cell *12*, 437-447.

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, J., Tresenrider, A., Chia, M., McSwiggen, D.T., Spedale, G., Jorgensen, V., Liao, H., van Werven, F.J., and Unal, E. (2017). Kinetochore inactivation by expression of a repressive mRNA. Elife *6*.

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, X., Yin, X., Li, J., Wu, Z., Qi, Y., Wang, X., Liu, W., and Xu, Y. (2021). Structures of the human Mediator and Mediator-bound preinitiation complex. Science *372*.

Cheng, Z., Otto, G.M., Powers, E.N., Keskin, A., Mertins, P., Carr, S.A., Jovanovic, M., and Brar, G.A. (2018). Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. Cell *172*, 910-923 e916.

Chia, M., Li, C., Marques, S., Pelechano, V., Luscombe, N.M., and van Werven, F.J. (2021). High-resolution analysis of cell-state transitions in yeast suggests widespread transcriptional tuning by alternative starts. Genome Biol *22*, 34.

Christensen, A.K., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am J Anat *178*, 1-10.

Compe, E., and Egly, J.M. (2012). TFIIH: when transcription met DNA repair. Nat Rev Mol Cell Biol *13*, 343-354.

Consortium, F., the, R.P., Clst, Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., *et al.* (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.

Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R., and Young, D.B. (2013). Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. Cell Rep *5*, 1121-1131.

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

Demircioglu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N.A., Kahles, A., Lehmann, K.V., Stegle, O., Brazma, A., *et al.* (2019). A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. Cell *178*, 1465-1477 e1417.

Deshpande, A.P., and Patel, S.S. (2012). Mechanism of transcription initiation by the yeast mitochondrial RNA polymerase. Biochim Biophys Acta *1819*, 930-938.

Devlin, C., Tice-Baldwin, K., Shore, D., and Arndt, K.T. (1991). RAP1 is required for BAS1/BAS2- and GCN4-dependent transcription of the yeast HIS4 gene. Mol Cell Biol *11*, 3642-3651.

Dhillon, N., Shelansky, R., Townshend, B., Jain, M., Boeger, H., Endy, D., and Kamakaka, R. (2020). Permutational analysis of Saccharomyces cerevisiae regulatory elements. Synth Biol (Oxf) *5*, ysaa007.

Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Mol Cell Biol *27*, 5575-5586.

Donahue, T.F., Daves, R.S., Lucchini, G., and Fink, G.R. (1983). A short nucleotide sequence required for regulation of HIS4 by the general control system of yeast. Cell *32*, 89-98.

Dujon, B. (1996). The yeast genome project: what did we learn? Trends Genet *12*, 263-270.

Dumont, M.E., Ernst, J.F., Hampsey, D.M., and Sherman, F. (1987). Identification and sequence of the gene encoding cytochrome c heme lyase in the yeast Saccharomyces cerevisiae. EMBO J *6*, 235-241.

Espinoza, F.H., Farrell, A., Nourse, J.L., Chamberlin, H.M., Gileadi, O., and Morgan, D.O. (1998). Cak1 is required for Kin28 phosphorylation and activation in vivo. Mol Cell Biol *18*, 6365-6373.

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. Molecular and Cellular Biology *21*, 4427-4440.

Faye, G., Simon, M., Valay, J.G., Fesquet, D., and Facca, C. (1997). Rig2, a RING finger protein that interacts with the Kin28/Ccl1 CTD kinase in yeast. Molecular and General Genetics *255*, 460-466.

Fazal, F.M., Meng, C.A., Murakami, K., Kornberg, R.D., and Block, S.M. (2015). Real-time observation of the initiation of RNA polymerase II transcription. Nature *525*, 274-277.

Feaver, W.J., Henry, N.L., Bushnell, D.A., Sayre, M.H., Brickner, J.H., Gileadi, O., and Kornberg, R.D. (1994). Yeast TFIIE. Cloning, expression, and homology to vertebrate proteins. J Biol Chem *269*, 27549-27553.

Feaver, W.J., Henry, N.L., Wang, Z., Wu, X., Svejstrup, J.Q., Bushnell, D.A., Friedberg, E.C., and Kornberg, R.D. (1997). Genes for Tfb2, Tfb3, and Tfb4 subunits of yeast transcription/repair factor IIH. Homology to human cyclin-dependent kinase activating kinase and IIH subunits. J Biol Chem *272*, 19319-19327.

Feaver, W.J., Huang, W., Gileadi, O., Myers, L., Gustafsson, C.M., Kornberg, R.D., and Friedberg, E.C. (2000). Subunit interactions in yeast transcription/repair factor TFIIH. Requirement for Tfb3 subunit in nucleotide excision repair. J Biol Chem *275*, 5941-5946.

Fisher, R.P., Jin, P., Chamberlin, H.M., and Morgan, D.O. (1995). Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase. Cell *83*, 47-57.

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.

Fujiwara, R., Damodaren, N., Wilusz, J.E., and Murakami, K. (2019). The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. Proc Natl Acad Sci U S A *116*, 22573-22582.

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.

Garavis, M., Gonzalez-Polo, N., Allepuz-Fuster, P., Louro, J.A., Fernandez-Tornero, C., and Calvo, O. (2017). Sub1 contacts the RNA polymerase II stalk to modulate mRNA synthesis. Nucleic Acids Res *45*, 2458-2471.

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.

Gervais, V., Busso, D., Wasielewski, E., Poterszman, A., Egly, J.M., Thierry, J.C., and Kieffer, B. (2001). Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family. J Biol Chem *276*, 7457-7464.

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. Mol Cell Biol *24*, 10975-10985.

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.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc *2*, 31-34.

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. P Natl Acad Sci USA *108*, 3566-3571.

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.

Gout, J.F., Li, W., Fritsch, C., Li, A., Haroon, S., Singh, L., Hua, D., Fazelinia, H., Smith, Z., Seeholzer, S., *et al.* (2017). The landscape of transcription errors in eukaryotic cells. Sci Adv *3*, e1701484.

Greber, B.J., Nguyen, T.H.D., Fang, J., Afonine, P.V., Adams, P.D., and Nogales, E. (2017). The cryo-electron microscopy structure of human transcription factor IIH. Nature.

Greber, B.J., Toso, D.B., Fang, J., and Nogales, E. (2019). The complete structure of the human TFIIH core complex. Elife *8*.

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.

Grunberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. Trends Biochem Sci *38*, 603-611.

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.

Guarente, L., Lalonde, B., Gifford, P., and Alani, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of S. cerevisiae. Cell *36*, 503-511.

Guarente, L., and Mason, T. (1983). Heme regulates transcription of the CYC1 gene of S. cerevisiae via an upstream activation site. Cell *32*, 1279-1286.

Hahn S, H.E., Guarente L (1985). Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of Saccharomy cescerevisiae. Proceedings of the National Academy of Sciences *82*, 8562-8566.

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

Hashimoto, S., Suzuki, Y., Kasai, Y., Morohoshi, K., Yamada, T., Sese, J., Morishita, S., Sugano, S., and Matsushima, K. (2004). 5'-end SAGE for the analysis of transcriptional start sites. Nat Biotechnol *22*, 1146-1149.

He, X., Khan, A.U., Cheng, H., Pappas, D.L., Jr., Hampsey, M., and Moore, C.L. (2003). Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. Genes Dev *17*, 1030-1042.

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

Healy, A.M., Helser, T.L., and Zitomer, R.S. (1987). Sequences required for transcriptional initiation of the Saccharomyces cerevisiae CYC7 genes. Mol Cell Biol *7*, 3785-3791.

Hein, P.P., Palangat, M., and Landick, R. (2011). RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. Biochemistry *50*, 7002-7014.

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.

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol *59*, 407-450.

Hochschild, A. (2015). Mastering Transcription: Multiplexed Analysis of Transcription Start Site Sequences. Mol Cell *60*, 829-831.

Holstege, F.C., Fiedler, U., and Timmers, H.T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. EMBO J *16*, 7468-7480.

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.

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.

Inostroza, J., Flores, O., and Reinberg, D. (1991). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE. J Biol Chem *266*, 9304-9308.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S.T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol *14*, 116-127.

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.

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.

Jona, G., Livi, L.L., and Gileadi, O. (2002). Mutations in the RING domain of TFB3, a subunit of yeast transcription factor IIH, reveal a role in cell cycle progression. J Biol Chem *277*, 39409-39416.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. Wiley Interdiscip Rev Dev Biol *1*, 40-51.

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.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

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.

Kaster, B.C., Knippa, K.C., Kaplan, C.D., and Peterson, D.O. (2016). RNA Polymerase II Trigger Loop Mobility: INDIRECT EFFECTS OF Rpb9. J Biol Chem *291*, 14883-14895.

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.

Kim, D., Hong, J.S.J., Qiu, Y., Nagarajan, H., Seo, J.H., Cho, B.K., Tsai, S.F., and Palsson, B.O. (2012). Comparative Analysis of Regulatory Elements between Escherichia coli and Klebsiella pneumoniae by Genome-Wide Transcription Start Site Profiling. Plos Genetics *8*.

Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599-608.

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.

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., Sumiya, E., Nagata, M., Ito, T., and Sekimizu, K. (2008). Transcriptional repression of the IMD2 gene mediated by the transcriptional co-activator Sub1. Genes Cells *13*, 1113-1126.

Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc Natl Acad Sci U S A *83*, 2850-2854.

Kuehner, J.N., and Brow, D.A. (2006). Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. J Biol Chem *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.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software *28*, 1-26.

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.

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.

Lis, C.G.a.J.T. (1993). DNA Melting on Yeast RNA Polymerase II Promoter. Science *261*, 759-762.

Liu, R., Liu, L., Li, X., Liu, D., and Yuan, Y. (2020). Engineering yeast artificial core promoter with designated base motifs. Microb Cell Fact *19*, 38.

Lu, Z., and Lin, Z. (2019). Pervasive and dynamic transcription initiation in Saccharomyces cerevisiae. Genome Res *29*, 1198-1210.

Lu, Z., and Lin, Z. (2020). The origin and evolution of a distinct mechanism of transcription initiation in yeasts. Genome Res.

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 Research *25*, 1008-1017.

Luo, J., Cimermancic, P., Viswanath, S., Ebmeier, C.C., Kim, B., Dehecq, M., Raman, V., Greenberg, C.H., Pellarin, R., Sali, A., *et al.* (2015). Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. Mol Cell *59*, 794-806.

Luse, D.S., Parida, M., Spector, B.M., Nilson, K.A., and Price, D.H. (2020). A unified view of the sequence and functional organization of the human RNA polymerase II promoter. Nucleic Acids Res *48*, 7767-7785.

Maicas, E., and Friesen, J.D. (1990). A sequence pattern that occurs at the transcription initiation region of yeast RNA polymerase II promoters. Nucleic Acids Res *18*, 3387-3393.

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*.

Malik, I., Qiu, C., Snavely, T., and Kaplan, C.D. (2017). Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo. Nucleic Acids Res *45*, 4431-4451.

Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev *8*, 515-524.

McNeil, J.B. (1988). Functional characterization of a pyrimidine-rich element in the 5'-noncoding region of the yeast iso-1-cytochrome c gene. Molecular and Cellular Biology.

McNeil, J.B., and Smith, M. (1985). Saccharomyces cerevisiae CYC1 mRNA 5'-end positioning: analysis by in vitro mutagenesis, using synthetic duplexes with random mismatch base pairs. Mol Cell Biol *5*, 3545-3551.

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.

Miura, F., Kawaguchi, N., Sese, J., Toyoda, A., Hattori, M., Morishita, S., and Ito, T. (2006). A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. Proc Natl Acad Sci U S A *103*, 17846-17851.

Murakami, K., Gibbons, B.J., Davis, R.E., Nagai, S., Liu, X., Robinson, P.J., Wu, T., Kaplan, C.D., and Kornberg, R.D. (2012). Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. Proc Natl Acad Sci U S A *109*, 4816-4821.

Murakami, K., Mattei, P.J., Davis, R.E., Jin, H., Kaplan, C.D., and Kornberg, R.D. (2015). Uncoupling Promoter Opening from Start-Site Scanning. Mol Cell *59*, 133-138.

Nagawa, F., and Fink, G.R. (1985). The Relationship between the Tata Sequence and Transcription Initiation Sites at the His4 Gene of Saccharomyces-Cerevisiae. P Natl Acad Sci USA *82*, 8557-8561.

Natsoulis, G., Hilger, F., and Fink, G.R. (1986). The HTS1 gene encodes both the cytoplasmic and mitochondrial histidine tRNA synthetases of S. cerevisiae. Cell *46*, 235-243.

Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M., Sheng, Y., Abdelhamid, R.F., Anand, S., *et al.* (2013). Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. Genome Res *23*, 1938-1950.

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.

Ohkuma, Y., Hashimoto, S., Wang, C.K., Horikoshi, M., and Roeder, R.G. (1995). Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha. Mol Cell Biol *15*, 4856-4866.

Okuda, M., Tanaka, A., Arai, Y., Satoh, M., Okamura, H., Nagadoi, A., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. (2004). A novel zinc finger structure in the large subunit of human general transcription factor TFIIE. J Biol Chem *279*, 51395-51403.

Okuda, M., Tanaka, A., Satoh, M., Mizuta, S., Takazawa, M., Ohkuma, Y., and Nishimura, Y. (2008). Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. EMBO J *27*, 1161-1171.

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.

Pardee, T.S., Bangur, C.S., and Ponticelli, A.S. (1998). The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. J Biol Chem *273*, 17859-17864.

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.

Peckham, H.E., Thurman, R.E., Fu, Y., Stamatoyannopoulos, J.A., Noble, W.S., Struhl, K., and Weng, Z. (2007). Nucleosome positioning signals in genomic DNA. Genome Res *17*, 1170-1177.

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

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., Wei-Hua Wu, Jong G. Na, and Michael Hampsey (1994). Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. J Biol Chem.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature *533*, 353-358.

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., *et al.* (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376-380.

Policastro, R.A., Raborn, R.T., Brendel, V.P., and Zentner, G.E. (2020). Simple and efficient profiling of transcription initiation and transcript levels with STRIPE-seq. Genome Res *30*, 910-923.

Qiu, C., Erinne, O.C., Dave, J.M., Cui, P., Jin, H., Muthukrishnan, N., Tang, L.K., Babu, S.G., Lam, K.C., Vandeventer, P.J., *et al.* (2016). High-Resolution Phenotypic Landscape of the RNA Polymerase II Trigger Loop. PLoS Genet *12*, e1006321.

Qiu, C., Jin, H., Vvedenskaya, I., Llenas, J.A., Zhao, T., Malik, I., Visbisky, A.M., Schwartz, S.L., Cui, P., Čabart, P., *et al.* (2020). Universal promoter scanning by Pol II during transcription initiation in Saccharomyces cerevisiae. Genome Biology *21*.

Rabut, G., Le Dez, G., Verma, R., Makhnevych, T., Knebel, A., Kurz, T., Boone, C., Deshaies, R.J., and Peter, M. (2011). The TFIIH subunit Tfb3 regulates cullin neddylation. Mol Cell *43*, 488-495.

Ranish, J.A., and Hahn, S. (1991). The Yeast General Transcription Factor Tfiia Is Composed of 2 Polypeptide Subunits. Journal of Biological Chemistry *266*, 19320-19327.

Ranish, J.A., Hahn, S., Lu, Y., Yi, E.C., Li, X.J., Eng, J., and Aebersold, R. (2004). Identification of TFB5, a new component of general transcription and DNA repair factor IIH. Nat Genet *36*, 707-713.

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

Riordan, D.P., Herschlag, D., and Brown, P.O. (2011). Identification of RNA recognition elements in the Saccharomyces cerevisiae transcriptome. Nucleic Acids Res *39*, 1501-1509.

Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell *166*, 1411-1422 e1416.

Roca, X., Olson, A.J., Rao, A.R., Enerly, E., Kristensen, V.N., Borresen-Dale, A.L., Andresen, B.S., Krainer, A.R., and Sachidanandam, R. (2008). Features of 5'-splice-site efficiency derived from disease-causing mutations and comparative genomics. Genome Res *18*, 77-87.

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.

Rolfes, R. (2006). Regulation of purine nucleotide biosynthesis: in yeast and beyond.

Ruohonen, L., Penttila, M., and Keranen, S. (1991). Optimization of Bacillus alpha-amylase production by Saccharomyces cerevisiae. Yeast *7*, 337-346.

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

Schilbach, S., Aibara, S., Dienemann, C., Grabbe, F., and Cramer, P. (2021). Structure of RNA polymerase II pre-initiation complex at 2.9 A defines initial DNA opening. Cell.

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., and Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature.

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.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol *19*, 65-71.

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.

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.

Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell *57*, 103-113.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res *27*, 491-499.

Strathern, J.N., Jones, E.W., and Broach, J.R. (1982). The Molecular biology of the yeast Saccharomyces : metabolism and gene expression (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. Annu Rev Biochem *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, 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.

Suzuki, Y., Taira, H., Tsunoda, T., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., *et al.* (2001). Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. EMBO Rep *2*, 388-393.

Svejstrup, J.Q., Feaver, W.J., LaPointe, J., and Kornberg, R.D. (1994). RNA polymerase transcription factor IIH holoenzyme from yeast. J Biol Chem *269*, 28044-28048.

Tareen, A., and Kinney, J.B. (2020). Logomaker: beautiful sequence logos in Python. Bioinformatics *36*, 2272-2274.

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. Mol Cell *31*, 671-682.

Thorsen, K., Schepeler, T., Oster, B., Rasmussen, M.H., Vang, S., Wang, K., Hansen, K.Q., Lamy, P., Pedersen, J.S., Eller, A., *et al.* (2011). Tumor-specific usage of alternative transcription start sites in colorectal cancer identified by genome-wide exon array analysis. BMC Genomics *12*, 505.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

Tsay, Y.H., and Robinson, G.W. (1991). Cloning and characterization of ERG8, an essential gene of Saccharomyces cerevisiae that encodes phosphomevalonate kinase. Mol Cell Biol *11*, 620-631.

van Eeuwen, T., Li, T., Kim, H.J., Gorbea Colon, J.J., Parker, M.I., Dunbrack, R.L., Garcia, B.A., Tsai, K.L., and Murakami, K. (2021). Structure of TFIIK for phosphorylation of CTD of RNA polymerase II. Sci Adv *7*.

Vo ngoc, L., Huang, C.Y., Cassidy, C.J., Medrano, C., and Kadonaga, J.T. (2020). Identification of the human DPR core promoter element using machine learning. Nature *585*, 459-463.

Vvedenskaya, I.O., Goldman, S.R., and Nickels, B.E. (2015a). Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5' ends. Methods Mol Biol *1276*, 211-228.

Vvedenskaya, I.O., Goldman, S.R., and Nickels, B.E. (2018). Analysis of Bacterial Transcription by "Massively Systematic Transcript End Readout," MASTER. Methods Enzymol *612*, 269-302.

Vvedenskaya, I.O., Vahedian-Movahed, H., Zhang, Y., Taylor, D.M., Ebright, R.H., and Nickels, B.E. (2016). Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. Proc Natl Acad Sci U S A *113*, E2899-2905.

Vvedenskaya, I.O., Zhang, Y., Goldman, S.R., Valenti, A., Visone, V., Taylor, D.M., Ebright, R.H., and Nickels, B.E. (2015b). Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. Mol Cell *60*, 953-965.

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. J Biol Chem *284*, 19601-19612.

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.

Winkelman, J.T., Vvedenskaya, I.O., Zhang, Y., Zhang, Y., Bird, J.G., Taylor, D.M., Gourse, R.L., Ebright, R.H., and Nickels, B.E. (2016). Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. Science *351*, 1090-1093.

Wong, M.S., Kinney, J.B., and Krainer, A.R. (2018). Quantitative Activity Profile and Context Dependence of All Human 5' Splice Sites. Mol Cell *71*, 1012-1026 e1013.

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.

Yamashita, R., Sathira, N.P., Kanai, A., Tanimoto, K., Arauchi, T., Tanaka, Y., Hashimoto, S., Sugano, S., Nakai, K., and Suzuki, Y. (2011). Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. Genome Res *21*, 775-789.

Yang, C., Fujiwara, R., Kim, H.J., Basnet, P., Zhu, Y., Gorbea Colon, J.J., Steimle, S., Garcia, B.A., Kaplan, C.D., and Murakami, K. (2022). Structural visualization of de novo transcription initiation by Saccharomyces cerevisiae RNA polymerase II. Mol Cell.

Yang, C., and Ponticelli, A.S. (2012). Evidence that RNA polymerase II and not TFIIB is responsible for the difference in transcription initiation patterns between Saccharomyces cerevisiae and Schizosaccharomyces pombe. Nucleic Acids Res *40*, 6495-6507.

Yarden, G., Elfakess, R., Gazit, K., and Dikstein, R. (2009). Characterization of sINR, a strict version of the Initiator core promoter element. Nucleic Acids Res *37*, 4234-4246.

Young, S.K., and Wek, R.C. (2016). Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. J Biol Chem *291*, 16927-16935.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614-620.

Zhang, P., Dimont, E., Ha, T., Swanson, D.J., Consortium, F., Hide, W., and Goldowitz, D. (2017). Relatively frequent switching of transcription start sites during cerebellar development. BMC Genomics *18*, 461.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. Science *338*, 1076-1080.

Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. Nucleic Acids Res *33*, 2838-2851.

Zhao, T., Vvedenskaya, I.O., Lai, W.K., Basu, S., Pugh, B.F., Nickels, B.E., and Kaplan, C.D. (2021). Ssl2/TFIIH function in transcription start site scanning by RNA Polymerase II in Saccharomyces cerevisiae. Elife *10*.

Zheng, H., Sun, L., Peng, W., Shen, Y., Wang, Y., Xu, B., Gu, W., Chen, S., Huang, Z., and Wang, S. (2011). Global identification of transcription start sites in the genome of Apis mellifera using 5'LongSAGE. J Exp Zool B Mol Dev Evol *316*, 500-514.