# MECHANISTIC INSIGHTS INTO THE FUNCTION OF SSL2/TFIIH IN RNA POLYMERASE II TRANSCRIPTION START SITE SCANNING

by

**Tingting Zhao**

B.S., Lanzhou University of Technology, Lanzhou, China, 2009.

M.S., Xiamen (Amoy) University, Xiamen, China, 2012.

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

**Tingting Zhao**

It was defended on

October 22, 2019

and approved by

Karen Arndt, PhD, Professor, Dept. of Biological Sciences

Andrea J. Berman, PhD, Associate Professor, Dept. of Biological Sciences

Jon P. Boyle, PhD, Associate Professor, Dept. of Biological Sciences

C. Joel McManus, PhD, Associate Professor, Dept. of Biological Sciences, Carnegie Mellon University

Thesis Advisor: Craig D. Kaplan, PhD, Associate Professor, Dept. of Biological Sciences

**MECHANISTIC INSIGHTS INTO THE FUNCTION OF SSL2/TFIIH IN RNA POLYMERASE II TRANSCRIPTION START SITE SCANNING**

Tingting Zhao, PhD

University of Pittsburgh, 2019

The initial step of RNA polymerase II (Pol II) transcription involves a large number of transcription factors and arises at multiple sites within most promoters. TFIIH is an essential, multisubunit transcription factor that assembles on promoter DNA with Pol II and five other general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE and TFIIF to form a pre-initiation complex (PIC) for basal transcription. During transcription initiation, TFIIH melts promoter DNA through the ATPase activity of its Ssl2 subunit. In the model eukaryote *Saccharomyces cerevisiae*, after DNA melting, Pol II scans downstream for usable transcription start sites (TSSs). TFIIH has been proposed as responsible for promoter scanning steps downstream of promoter melting but this has not been tested extensively. *ssl2* mutations affect TSS selection, consistent with Ssl2-dependent functions of TFIIH in scanning. We hypothesize that TFIIH serves as the engine for scanning by affecting TSS usage in at least two possible ways: by controlling the processivity (how far) of scanning, and/or the rate at which scanning translocates (how fast). To understand the function of Ssl2/TFIIH in promoter scanning and TSS selection, we identified novel alleles of *SSL2* in genetic screens, mutants defective in TSS distribution that may potentially arise from altered scanning. Consistent with this notion, these *ssl2* alleles alter scanning in ways that are distinct from how changes to the Pol II active site alter scanning and this difference is observed genome-wide. To understand further how Ssl2/TFIIH and Pol II or other initiation factors work concurrently to promote transcription initiation, we performed genetic interaction

experiments between initiation factors. Our genetic data indicate that there are at least two major networks controlling promoter scanning and TSS selection, one controls the efficiency of initiation through Pol II activity or factors regulating Pol II's activity; another network hypothetically controls the processivity of scanning by TFIIH. Moreover, we are examining if promoter scanning is a conserved mechanism across eukaryotes. We are asking if perturbation of transcription initiation in *Schizosaccharomyces pombe* and *Drosophila melanogaster* through mutation of initiation factors fits expectations of a scanning mechanism. Our preliminary data in both organisms indicate differences from *S. cerevisiae*.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

## NOMENCLATURE

| | |
|---|---|
| Pol II | RNA polymerase II |
| GTF | General Transcription Factor |
| TFII | Transcription Factor II |
| PIC | Pre-Initiation Complex |
| DNA | Deoxyribonucleic Acid |
| RNA | Ribonucleic Acid |
| cDNA | complementary DNA |
| TSS | Transcription Start Site |
| TBP | TATA Binding Protein |
| CC | Closed complex |
| OC | Open complex |
| CP | Core Promoter |
| CPE | Core Promoter Element |
| bp | Basepair |
| nt | Nucleotide |
| Inr | Initiator |
| BRE | TFIIB recognition elements |
| DPE | Downstream core promoter element |
| TAF | TBP-associated factor |
| A | Adenine |
| T | Thymine |

| | |
|---|---|
| G | Guanine |
| C | Cytosine |
| Y | Pyrimidine, Cytosine/Thymine |
| R | Purine, Adenine/Guanine |
| W | A/T, Adenine/Thymine |
| K | T/G, Thymine/Guanine |
| CpG | Cytosine-phosphodiester bond-Guanine |
| SAGA | Spt-Ada-Gcn5-acetyltransferase |
| SWI/SNF | SWItch/Sucrose Non-Fermentable |
| ATP | Adenosine triphosphate |
| ITC | Initial transcription complex |
| EC | Elongation complex |
| NTD | N terminal domain |
| CTD | C terminal domain |
| NER | Nucleotide excision repair |
| WT | Wild-type |
| SF | Super family |
| ssDNA | Single strand DNA |
| dsDNA | Double strand DNA |
| HD | Helicase domain |
| DRD | Damage recognition domain |
| ThM | Thumb-like motif |
| C-ter | C terminal extension |

| | |
|---|---|
| MPA | Mycophenolic Acid |
| NFR | Nucleosome free region |
| TL | Trigger loop |
| GOF | Gain of function |
| LOF | Loss of function |

# PREFACE

During my graduate education I have grown both scientifically and mentally. The progress I have made in science and my personal life over the past six years could have not happened without the support of many people. I wish I could thank all of them but space limited to only express my thankfulness to some.

First I would like to thank my PhD thesis advisor Dr. Craig D. Kaplan for the great influence he has had on me, both in science and in my personal life. Dr. Kaplan is the person I have spent the most time with in the past six years. Although there is not enough space to list all the things I have learned from him, I have picked the top ten: problem-solving driven thinking; always add controls when you are solving problems; never get disappointed about unexpected results; comment on every document word by word; love mentoring; lead students to think instead of telling them what to do; always online; never let the high standard slide down; plan ahead and ask brilliant questions. These are things that I'm sure will benefit me for the rest of my career.

I would like to thank my committee members Dr. Hays Rye, Dr. Michel Polymenis, Dr. Junyuan Ji from Texas A&M university, and Dr. Karen Arndt, Dr. Jon Boyle, Dr. Andrea Berman from University of Pittsburgh, and Dr. Joel McManus from Carnegie Mellon University, for their guidance in every discussion of my research. I would like to thank Dr. Mary Bryk, Dr. David Peterson, Dr. Sarah Hainer and Dr. Miler Lee for their great comments in every joint group meeting. I would like thank Jun Fan from China Agricultural University for his encouragement in continuing my study in science. Dr. Ping He and his research from Texas A&M University for letting me know the Biochemistry & Biophysics graduate program. Dr. Xiuren Zhang for great

advice in selecting a lab. Dr. Jennifer Herman, Dr. Dorothy Shippen and Dr. John Mullet for their mentoring and appreciation.

I would like to thank my current and past colleagues from the Kaplan lab. I would like to thank the previous lab mate Huiyan Jin for her mentoring during my early years in the Kaplan lab, including teaching me the first primer extension gel, and the extended talks, emails and messages after her graduation to answer both scientific and non-scientific questions. I would also like to thank my previous lab mates Indranil Malik and Chenxi Qiu for their guidance and support both during their stay in the Kaplan lab and after moved on. I appreciate having them as my colleagues, their hard-working attitudes always motivates me to work hard. I would also like to thank my current lab mates Anandhakumar Jayamani for all the time support, and my other lab mates Yunye Zhu, Payal Arora, Bingbing Duan, Ping Cui, Evelyn Callaway, Shrabani Basu and Alex Visbisky for their help for being excellent colleagues.

I would like to thank all the friends I've met after entering the US. My best friends Limeng Xie and Jisun Lee, for their tremendous support ever since the first time we met. I would like thank Xuan Lin for impacting me by bringing her scientific mind to daily. I appreciate the collaboration with Ji lab in the *Drosophila* program, where I met Xiaojun Xie, Yani Zheng, Xiao Li and Mengmeng Liu, and became close friends with Mengmeng. I would like to thank friends I met in Pittsburgh as well, Mitchell Ellison, Brendan McShane, Katherine Wozniak, Yang Liu, Zhihao Sun, Sarah Smith, Donya Shodja and Eden Wellesley McQueen for their help on setting up in Pittsburgh. I want to acknowledge Sena Zheng as a great roommate and well-timed friend, En Yu as a great gym partner, Lu Jiang as a friend and peer, Chenyao Lou and Yuesong Zou as friends and great coding mentors. Lastly, I would like to thank Bin Shen, Chonghua Li, Jianan Arthur Li and Nana Hong for their help during the time I was looking for a job.

I would like to thank my family members for their unconditional love and support. My mom was born in a small village as an eldest daughter in 1960's, where and when girls were not encouraged to go to school. She stole ten yuan, approximately 1.6 dollars, from my grandfather's pocket and registered for primary school when she was nine years old. That's how she started her education. When she graduated from high school, she was one of only three female students in the school. Her view on the value of education and her courage in breaking prejudice is the model I will chase for my entire life. My dad's support is everywhere. When I was in Lan Zhou, Xiamen, Beijing, US, he has always brought me anything he thought I might need. When I was struggling with my graduate student life, he resigned his job and came to accompany me for six months from thousands of miles away. When I needed someone to talk with face to face, he sent my cousin to me. He once said that I'm like a flying kite and the line is always in his hand, whenever I need him to, he can get close to me. My brother and my best friend, Taotao Zhao, we talk to each other almost every day. No words can express how much I appreciate having him in my life. Lastly, I would like to thank my cousins Zuoming Zhao and Zuozhou Zhao. They are always my role models, their advice and support since I decided to apply for graduate school and during my time in graduate school helped to get where I am today. Their quality as good family members and great scientists is like a lighthouse in my life to give me direction and strength. I appreciate having them in our big family and as role models and I hope to be a role model like them for our younger generations.

## DEDICATION

I dedicate this dissertation to my grandmother Huiying Pei, who passed away in May, 2015. She is the most elegant and generous lady I've ever met. Her wisdom in connecting our big family as a whole and the warm family union moment she created will be my lifetime spiritual nutrient.

# 1.0 INTRODUCTION

Transcription of eukaryotic protein-coding genes is carried out by RNA polymerase II (Pol II) in three sequential steps: initiation, elongation and termination. Accurate initiation requires minimally the assistance of five general transcription factors (GTFs) TFIIB, TFIID, TFIIE, TFIIF and TFIIH, which together with Pol II, comprise the basal transcription machinery. At the beginning of transcription, the basal transcription machinery assembles at a defined DNA region for each transcript called a promoter, melts the double-stranded DNA and selects a start site to initiate transcription. Ssl2, a subunit of the GTF TFIIH, and its function in TSS selection is the focus of this thesis. The majority of our work examines TSS selection in the model eukaryote *Saccharomyces cerevisiae*, however we additionally ask if mechanisms of initiation are conserved among eukaryotes. The first part of my thesis (Chapter II-IV) examines the function of Ssl2 in transcription initiation and TSS selection using budding yeast *Saccharomyces cerevisiae*, and the second part (Appendix A) investigates the conservation of the TSS selection mechanisms in fission yeast *Schizosaccharomyces pombe* and the fruit fly *Drosophila melanogaster*.

Here, I will first review the process of Pol II initiation and TSS selection, and compare promoter architectures and TSS usage among well-studied eukaryotes. Second, I will review the mechanism of TSS selection in *Saccharomyces cerevisiae*. Third, I will review the functions of Ssl2. Finally, I will review the other known transcription factors involved in TSS selection beyond Pol II and Ssl2.

## 1.1 POI II TRANSCRIPTION INITIATION AND THE PROMOTER ARCHITECTURE

### 1.1.1 Overview of Pol II transcription initiation

The assembly of the pre-initiation complex (PIC) is the first step of RNA Polymerase II (Pol II) transcription whereby Pol II and general transcription factors (GTFs) TFIIA, TFIIIB, TFIID, TFIIE, TFIIF and TFIIH are recruited to double-stranded promoter DNA[1]. In the classical PIC assembly model, TATA box-binding protein (TBP), a subunit of TFIID, recognizes and binds to promoter DNA[1]. This binding is stabilized by addition of TFIIA and followed by the recruitment of TFIIB, which in turn recruits Pol II and TFIIF to the complex. Finally, binding of TFIIE and TFIIH completes PIC assembly. The PIC then transitions from the closed complex (CC) state to the open complex (OC) state, in which the double-stranded DNA is unwound and the DNA template strand is positioned in the active center of Pol II[2,3]. This transition is promoted by a DNA translocase subunit of TFIIH, which in *S. cerevisiae* is the Ssl2 protein, in an ATP-dependent manner[4,5]. Previous studies suggest that Ssl2 binds at the downstream DNA of PIC and translocates on the DNA away from the rest of PIC components, yet being bound to the PIC, this walk translocates DNA in the opposite direction and pushes it into the active site of Pol II[5,6] **(Figure 1, PDB 5OF4)**[7]. The Pol II active site then locates a transcription start site (TSS) to initiate transcription.

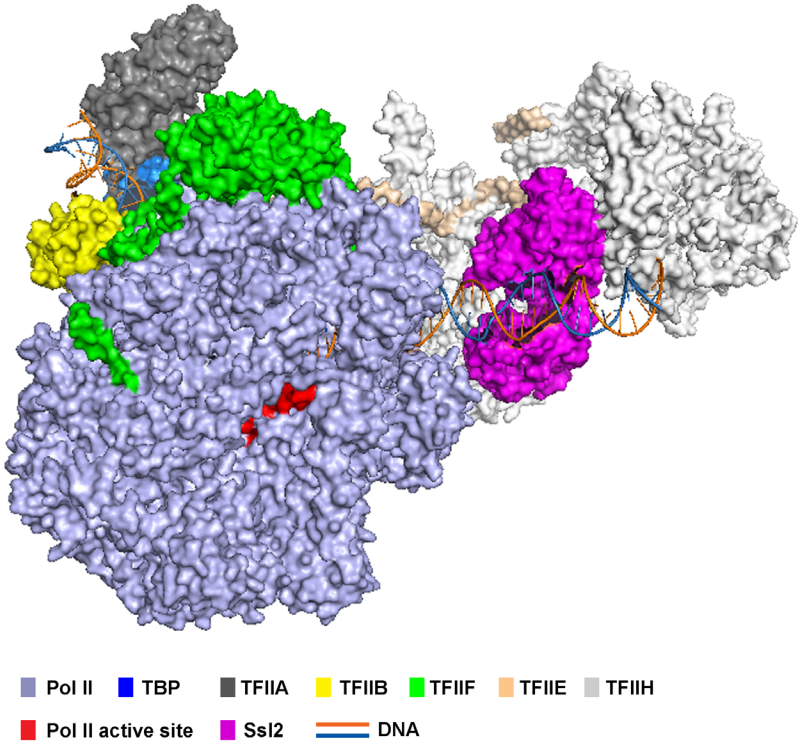



**Figure 1 Ssl2 assembles downstream inside of the PIC**

Shown is the cyo-EM structure of yeast transcription pre-initiation complex with TFIIH (PDB: 5OQJ)[7]. Ssl2 is highlighted (purple) within TFIIH complex to show its relative location to other PIC components.

### 1.1.2 Core promoters and core promoter elements

The minimal promoter sequence that is required to guide PIC assembly and ensure accurate transcription initiation is referred to as the core promoter (CP)[8,9]. In metazoans, a typical CP contains DNA sequences expanding from about 40 basepairs (bp) upstream to 40 bp downstream of the dominant TSS of a cluster (-40 to + 40 of TSS)[10]. Functional DNA motifs identified within CPs that are critical to the recruitment of basal transcription machinery and activation of Pol II transcription are designated as core promoter elements (CPEs). The best-known CPEs in metazoans are TATA-box, initiator (Inr), TFIIB recognition elements (BRE) and the downstream

core promoter element (DPE) **(Figure 2)**. **(1)** TATA box is the most famous CPE that contains an 8 bp consensus sequence (generally considered TATAWAAR (W=A/T, R=A/G) in *metazoans and* TATAWAWR in *S. cerevisiae*) and provides a binding site for TBP to nucleate PIC assembly[9,11]. Both the TATA box and the TBP are widely conserved from archaebacteria to humans[9,12]. However, there is no universal CPE for all promoters and TATA-box containing promoters represent only about 10-20% of the most eukaryotic promoters[9], while some organisms were found to lack TATA-boxes almost entirely, like *Toxoplasma gondii*[13]. Therefore, the majority promoters are classified as TATA-less promoters. **(2)** The Inr is a CPE that encompasses a TSS and surrounding bases, which is proposed to be recognized by TAF subunits of TFIID to direct PIC assembly in the absence of the TATA box[14]. The consensus sequences for Inr have been identified in humans as $YYA_{+1}NWYY$ (Y=C/T, N=any base, +1 is dominant TSS) and in *Drosophila* as $TCA_{+1}KTY$ (K=T/G) through measuring TFIID binding or transcriptional activity[9,15,16]. However, genome-wide mapping of the human TSS indicates consensus Inr of $Y_{-1}R_{+1}$ (-1 is one nucleotide upstream of +1 TSS)[17,18]. The prevalence of YR at the -1 and +1 positions of the non-template strand is found in both human and yeast Pol II promoters[19,20] (and our data in Chapter IV). This YR preference was explained by the biochemical stacking interactions between the two purines, one on the template -1 position and the other being the incoming preferred ATP or GTP for the +1 position of the RNA[21]. **(3)** The BRE is the DNA motif adjacent to TATA box and directly bound by TFIIB in a sequence dependent manner. When tested in different experimental systems, the BRE could have either a positive or negative effect on basal transcription activity[9,22-24]. For example, a transcription system using human cell extract showed that the BRE sequence was recognized by TFIIB and was found to stabilize TFIIB-TBP-promoter complex, thus supporting transcription initiation[22]. However, another study with both crude human nuclear extract and living

cells independently showed that BRE acted as a suppressor of basal transcription[23]. **(4)** DPE, found in both TATA-box containing and TATA-less promoters, was initially identified and commonly found in *Drosophila* TATA-less promoters[9]. The DPE is located downstream of Inr and is also recognized and bound by TFIID subunits TAFs[25]. The space between TSS and DPE was found critical for TFIID's binding affinity and the basal transcription level[26]. Inr and DPE often work as a combination for regulating transcription initiation[9].

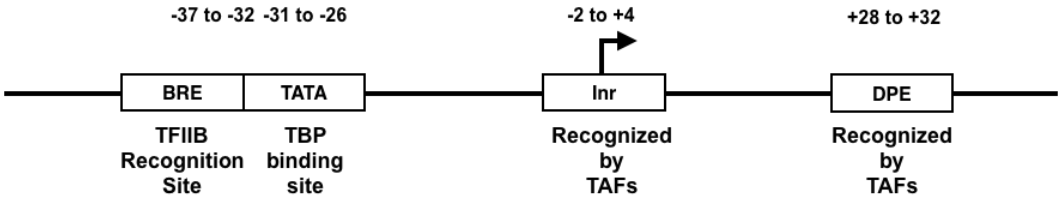



**Figure 2 Structure of core promoter in mammalian genes**

Adapted from Jennifer E.F. Butler *et.al*[8]. CPEs TATA, BRE, Inr and DPE in mammalian promoters are shown in boxes. The relative location of CPE to the transcription start site (+1) is labeled on top and the it's association factor during PIC assembly is shown on bottom.

### 1.1.3 Focused and dispersed transcription

Pol II transcription initiation has been classified into two patterns in terms of TSS distributions: focused or dispersed transcription **(Figure 3)**[9]. Focused initiation refers to the usage of a single TSS or a cluster of TSSs in a small region, ~10 bp, whereas dispersed transcription uses multiple TSSs that may be distributed over a 100 bp promoter region. An early study in mouse cells observed dispersed transcription at housekeeping genes, for example, at the HMG-CoA reductase promoter[9,27]. In contrast, focused transcription was thought to correlate with regulated or inducible genes, for example, *Drosophila* genes with restricted spatial expression patterns[9,28].

Studies in mammals found that TATA boxes were strongly overrepresented in focused (peaked) promoters, whereas CpG islands were highly associated with dispersed (broad) promoters[17]. Although, the linkage between CpG islands and the dispersed transcription pattern does not exist in *Drosophila*, likely due to low level of CpG methylation and a more even distribution of CpG dinucleotides across the *Drosophila* genome, *Drosophila* also adopts two shapes of TSS distributions, focused or dispersed[9]. Similar to the relationship between the CPE and the TSS distribution observed in mammals, *Drosophila* TATA, Inr and DPE are enriched in focused promoters[28]. Additionally, there is a good correlation in *Drosophila* between focused promoters and more tissue-restricted gene expression; in contrast, dispersed promoters are highly associated with ubiquitously expressed promoters[28].

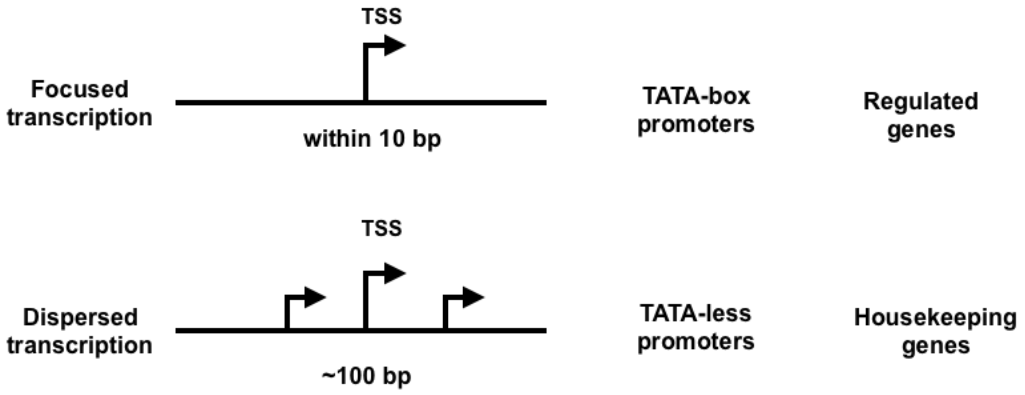



**Figure 3 Two patterns of transcription initiation and their relevant CPEs in mammalian genes**

Adapted from James T. Kadonaga[9]. Focused transcription initiation use a single TSS or tightly clustered TSSs within 10 bp promoter region and often found in regulated genes with a TATA-box. Dispersed transcription initiation use multiple TSSs distributed around 100 bp promoter region and ustually found in housekeeping genes associated with TATA-less promoters.

### 1.1.4 Yeast TATA transcription *vs.* metazoan TATA transcription

Similar to metazoans, only about 20% of yeast promoters contain a TATA-box and multiple TSSs are broadly used across most promoters[10,11]. However, unlike mammalian TATA-box containing promoters, which use a single TSS at 30 or 31 bp downstream, yeast TATA-box containing promoters use multiple TSSs that are distributed at 40-120 bp downstream[29-31] **(Figure 4)**.

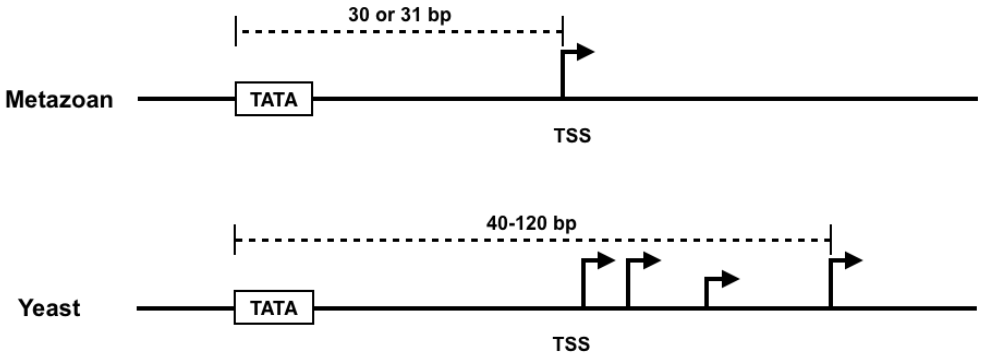



**Figure 4 Comparison of transcription initiation between metazoan and yeast TATA promoters**

Metazoan TATA-box containing promoters use a single TSS 30 or 31 bp downstream, whereas yeast TATA promoters used multiple TSSs distributed 40-120 bp downstream .

### 1.1.5 TFIID-dependent *vs.* SAGA-dependent transcription

Promoters in yeast have additionally been divided into TFIID- or SAGA (Spt-Ada-Gcn5-acetyltransferase)-dominated classes based on how TBP is hypothesized to be delivered to the core promoter[32-34]. In yeast, TFIID is composed of TBP and 14 TBP-associated factors (TAFs), Taf1-Taf14. Like TBP, Taf1-Taf13 are conserved from yeast to humans, whereas Taf14 is yeast specific, which also serves as subunit of TFIIF and the INO80 and SWI/SNF chromatin remodeler

complexes[34]. The largest subunit of TFIID, Taf1, contains a TBP binding domain that is critical to deliver TBP to promoters lacking a TATA-box[35]. Other TAFs in higher eukaryotes interact with Inr or DPE for promoter recognition[36]. Genome-wide gene expression analysis indicates that 90% of *S. cerevisiae* Pol II regulated genes rely on TFIID, mostly at TATA-less promoters[37]. Consistently, studies examining the dependency of TAFs on gene activation reveal that TATA-less promoters show a higher dependency on Taf1[37]. TBP could alternatively be delivered via SAGA complex, usually at TATA-containing promoters[33]. The SAGA complex contains 20 subunits and shares some TAFs with TFIID. Gene expression analysis and TAF-dependency experiments show that 10% of Pol II regulated genes rely on SAGA and are depleted of Taf1 (summarized in **Table 1**). Contrary to what was previously suggested, two recent studies indicate that TFIID and SAGA are both required for all yeast promoters[38,39].

**Table 1 Architecture of TFIID- or SAGA-dominated gene expression**

| Coactivator | TATA status | Taf1 dependency | Genome composition |
|---|---|---|---|
| TFIID | TATA-less | Taf1-donimated | ~90% |
| SAGA | TATA | Taf1-depleted | ~10% |

## 1.2 OPEN COMPLEX FORMATION AND TRANSCRIPTION START SITE SCANNING

### 1.2.1 DNA melting in metazoans *vs.* in *Saccharomyces cerevisiae*

During transcription initiation, double stranded DNA within the initial PIC needs to be melted to allow Pol II access for transcription. This DNA opening process is dependent on TFIIH

subunit XPB/Ssl2's ATPase activity[4,40-42]. Further study in *S. cerevisiae* revealed the ATP-dependent translocase mechanism for Ssl2 in DNA opening and OC formation[5]. This later study suggested a model that Ssl2 assembles at downstream site of PIC and tracks away from the PIC, but because it is still bound to other PIC components, its tracking motion is converted into an action that opens and inserts DNA into the Pol II active site. While *S. cerevisiae* and higher eukaryotes both show melted promoter DNA around 20 bp downstream of the TATA-box, transcription initiates almost immediately from where DNA melting occurs in other eukaryotes but not in *S. cerevisiae*[40,43]. In most studied eukaryotes for TATA-promoters, transcription initiates ~10 bp downstream from the initially melted region, with a ~15 bp transcription bubble that spans from -12 to +3 position relative to the TSS **(Figure 5)**[40]. In contrast, *S. cerevisiae* initiates at a region that is ~20-100 bp further downstream from where the DNA is initially melted[43]. This was demonstrated for the *S. cerevisiae GAL1* and *GAL10* promoters, with TATA-TSS distances of 84 bp or 114 bp respectively, and showed that DNA melting started 20 bp downstream of TATA and extended through the TSS region.

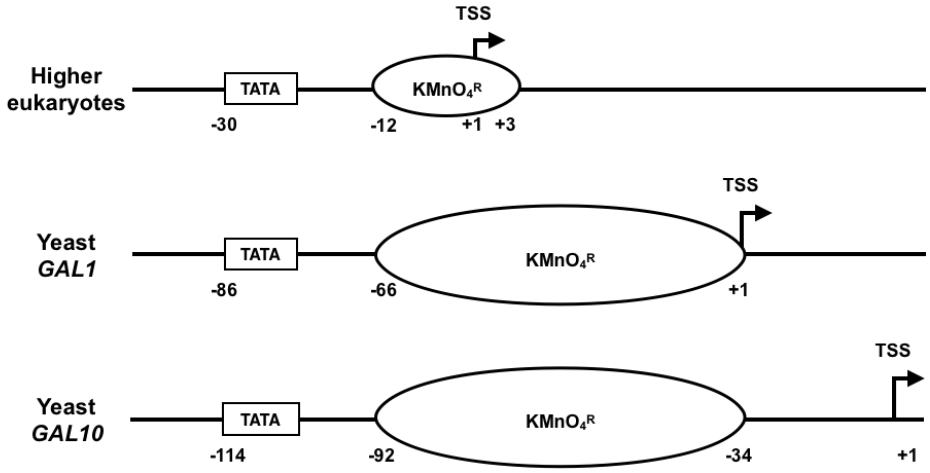



**Figure 5 DNA melting at other studied eukaryotic or at *S. cerevisiae* Pol II promoters**

DNA melting occurs ~20 bp downstream of TATA element in both metazoan and yeast promoters. However, TSS is closer in metazoan TATA-containing promoter than it is in *S. cerevisiae* promoter, ~30 bp *vs.* ~40-120 bp

downstream of TATA. Promoter melting is measured by "the reactivity to $KMnO_4$ ($KMnO_4^R$)": briefly, thymine on single stranded DNA, representing an opened promoter state, could be modified by $KMnO_4$. The modified site could be cleaved by adding piperidine and viewed by the ligation-mediated polymerase chain reaction (LMPCR) as described[43]. In higher eukaryotes, promoter melting extends from 20 bp downstream of TATA to +3 relative to TSS. In *S. cerevisiae*, promoter DNA melts from 20 bp downstream of TATA to TSS or -34 relative to TSS in *GAL1* and *GAL10*, respectively[40,43]. The architecture of the transcription bubble inside of the PIC is unknown, here, it is schematically showed as a bubble.

### 1.2.2 Evidence for promoter scanning in *S. cerevisiae*

The observation of TSSs occurring downstream of where DNA melting starts in *S. cerevisiae* led to the proposal for a scanning mechanism for TSS selection: in this model, Pol II and GTFs bind to the upstream promoter region and scan downstream for usable TSSs[43]. A critical study demonstrated this by using duplicated Inr regions in *S. cerevisiae*. This study demonstrated that an upstream Inr is preferred to a downstream one even when they are identical; however, when the upstream Inr was mutagenized, the downstream Inr became highly efficient. This observation supported that Pol II was capable of scanning by demonstration of polar preference for TSSs **(Figure 6)**[20].

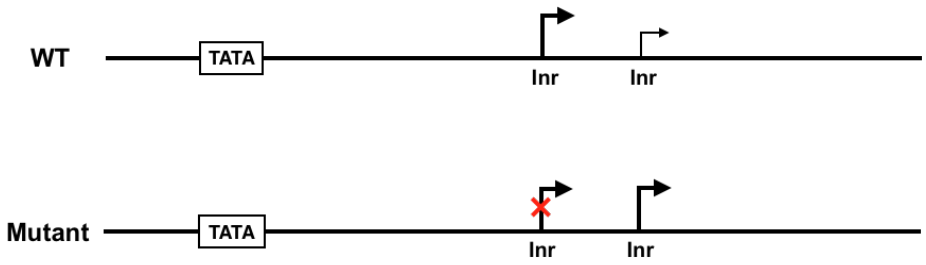



**Figure 6 Scanning is supported by the increased Inr usage at downstream site when the identical upstream one is mutagenized**

In the scanning model, Pol II is hypothesized to bind upstream of promoter and scans to downstream for TSS unidirectionally. When one of the identical Inr at upstream site is mutagenized, initiation at downstream one is increased.

A recent single molecule study using an *in vitro* reconstituted *S. cerevisiae* transcription assay showed that after initial ~15 bp transcription bubble formation, the distance between the PIC assembly site and the downstream DNA was further shortened ~85 bp on average, indicating movement of downstream DNA toward the PIC by ~85 bp[44]. The architecture of the transcription bubble over this putative scanning region downstream is unclear, and has been suggested to exist either as a large extended bubble or a small bubble that translocates **(Figure 7)**. This shortening of the distance between the PIC and a marker of downstream DNA was driven a TFIIH's ATPase activity, presumably Ssl2's translocase activity (and not Rad3, consistent with all previous studies showing Ssl2 and not Rad3 functions in initiation), and is attributed to TSS scanning[5].

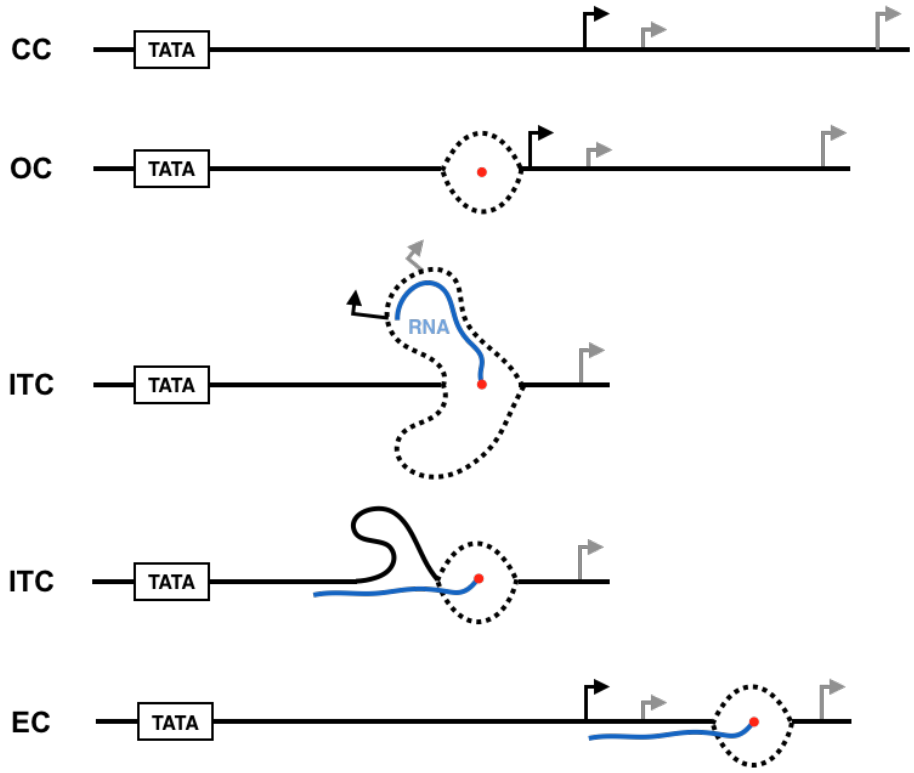



**Figure 7 Downstream DNA is drawn toward the PIC during *S. cerevisiae* promoter scanning**

Adapted from Furqan M. Fazal *et. al*[44]. Double stranded DNA within the closed complex (CC) of transcription machinery is melted to form an opening complex (OC), followed by additional DNA melting trough a large bubble extension or a small bubble translication and the initial transcription complex (ITC) formation. The extended bubble then collapse back to a closed or open state, followed by continuous downstream movement and the elongation complex (EC) formation. Shown is an example of the first TSS (black arrow) being used among other TSSs (grey arrows) by Pol II active site (red dot) and generating an RNA transcript (blue line). The bubble configuration inside the PIC is unclear (dashed bubble).

## 1.3 TFIIH AND ITS FUNCTIONS IN POL II TRANSCRIPTION INITIATION

### 1.3.1 TFIIH and its subunits

In *S. cerevisiae*, TFIIH is the largest general transcription factor and is composed of ten conserved subunits, Ssl2, Rad3, Tfb1, Tfb2, Ssl1, Tfb4, Tfb5, Ccl1, Tfb3, Kin28 and a yeast-specific subunit of Tfb6[3,45-49]. Human TFIIH subunits, listed with their yeast homologs are: XPB/Ssl2, XPD/Rad3, p62/Tfb1, p52/Tfb2, p44/Ssl1, p34/Tfb4, p8/Tfb5, cyclin H/Ccl1, MAT1/Tfb3 and CDK7/Kin28 **(Figure 8)**. TFIIH can be isolated both as a holoenzyme containing all subunits and as a core sub-complex that contains seven of the total subunits Rad3, Ssl2, Tfb1, Tfb2, Ssl1, Tfb4, and Tfb5[45,50,51]. The other three subunits absent in the core, Tfb3, Ccl1, and Kin28, form a kinase module, also known as TFIIK. The holoenzyme, core, and TFIIK can be differentiated by their distinct functions. The core is required for both transcription and nucleotide excision repair (NER), whereas the kinase module functions in transcription and cell cycle pathways. Among ten TFIIH subunits, three of them contain ATP-dependent catalytic activities: Kin28, which phosphorylates the C-terminal domain (CTD) of Pol II during transcription and promotes Pol II escape from the initiation complex; Rad3, a 5' to 3' DNA helicase that functions in DNA opening around the lesion during NER, and Ssl2, both an ATPase that functions in NER together with Rad3 and an ATP-dependent nucleic acid translocase that functions in DNA unwinding during transcription OC formation.

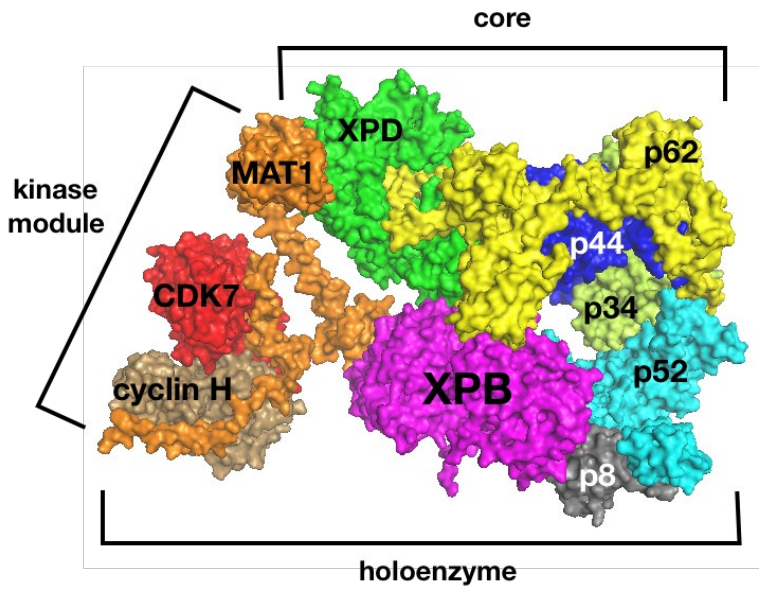



**Figure 8 Architecture of TFIIH**

As TFIIH structures from *S. cerevisiae* are missing some subunits, shown is the human TFIIH structure from Greber *et al* (PDB 6O9L)[52]. TFIIH is composed of XPB/Ssl2 (human/yeast names), XPD/Rad3, p62/Tfb1, p52/Tfb2, p44/Ssl1, p34/Tfb4, p8/Tfb5, cyclin H/Ccl1, MAT1/Tfb3 and CDK7/Kin28. Core sub-complex comprising XPB/Ssl2, XPD/Rad3, p62/Tfb1, p52/Tfb2, p44/Ssl1, p34/Tfb4 and p8/Tfb5, functions in both transcription and DNA repair. Kinase module includes cyclin H/Ccl1, MAT1/Tfb3 and CDK7/Kin28, and functions in transcription and cell cycle pathways.

### 1.3.2 Ssl2 in promoter opening and TSS scanning

The largest subunit of TFIIH, Ssl2, was originally identified in yeast as a suppressor of a stem-loop structure at the 5' untranslated region of the *HIS4* gene with a predicted role in transcription initiation[53], and, as a homologous gene of human *XPBC/ERCC3*, which functions in UV sensitivity and DNA repair[54]. The following research on XPB, the human homolog of Ssl2, demonstrated that ATPase activity of XPB is essential for transcription promoter opening[55]. Further evidence for Ssl2 function in transcription initiation came from mapping the RecA-like

domains of Ssl2 to the PIC, which suggested that Ssl2 promotes DNA melting by inserting ~15 bp of DNA into the Pol II active site[48]. The latter has been demonstrated by a single molecule experiment showing that the distance between the upstream DNA where the PIC assembles and the downstream DNA beyond TSS is shortened by an ATP-induced activity of PIC **(Figure 7)**[44]. This ATP-induced PIC activity for promoter opening has been proposed to be facilitated by an *in vitro*-demonstrated Ssl2 translocase activity within purified TFIIH[5]. Evidence of Ssl2's function in TSS selection comes from a study in *S. cerevisiae* showing some *ssl2* alleles shift start site distribution relative to the wild-type (WT) distribution[56]. These polar effects of *ssl2* mutants on TSS distributions are consistent with the "scanning model" where Pol II and GTFs scan from the upstream promoter to downstream positions for usable TSS[57].

### 1.3.3 Ssl2 and its enzymatic activities

Helicases are enzymes that use ATP hydrolysis to unwind double stranded nucleic acid into single strands[58]. They are classified into six families, SF1-SF6, based on nine short conserved amino-acid sequence motifs, Q, I, Ia, Ib, II, III, IV, V and VI, these motifs are designated as helicase motifs **(Table 2)**[58,59]. A newly proposed classification method suggests to additionally differentiate helicases into type A or B, based on their template nucleic acid as ssDNA or dsDNA, respectively; or, into type α or β, rely on the directionality of enzyme movement on template from 3′ to 5′ or from 5′ to 3′, separately[58,60]. SF1 and SF2 are the largest two helicase families, within which, each helicase contains at least seven helicase motifs, I, Ia, II–VI[60]. These helicase motifs are arranged within two core domains of a SF1 or SF2 helicase. The core domain is a helicase region that structurally forms a recombination protein RecA-like fold and functionally involved in NTP binding and hydrolysis[58]. Notably, among six families of helicases, only SF1 and SF2 family

members contain two core domains from a single polypeptide, whereas each helicase polypeptide in SF3-SF6 family contains only one core domain. Generally, SF1 and SF2 families contain both helicases of type A and B, whereas in terms of directionality, SF1 tends to be type α and SF2 has both type α and β[58]. Although enzymes that unwind double stranded nucleic acid are together defined as helicases and further classified into groups based on characteristics like template type or polarity, the mechanisms they use for unwinding are still significantly different[61]. Some helicases' couple unwinding with translocation, whereas others do not[61-63]. A subgroup of helicases that couple unwinding and translocation are referred to as translocases, which by definition are enzymes that use ATP hydrolysis to directionally translocate along single or double stranded nucleic acid[5,58]. Ssl2's activity in facilitating promoter DNA opening was originally identified as an SF2 family helicase activity, which hydrolyzes ATP to melt double stranded nucleic acid into single strands[55,58]. However, later studies indicate that Ssl2 possesses a 5' to 3' translocase activity during transcription initiation, which uses ATP hydrolysis to translocate along one strand of the double helix, because of a relatively fixed position in the PIC, Ssl2's translocase activity creates torsional stress that results in DNA melting [5,58].

**Table 2 Classification and characteristics of helicases**

| Classfier | Subtypes or attributes |
|---|---|
| Family | SF1, SF2, SF3, SF4, SF5, SF6 |
| Helicase motif | Q, I, Ia, Ib, II, III, IV, V, VI |
| Core domain | One or two RecA-like fold |
| Nucleic acid template | Single stranded (type A), double stranded (type B) |
| Directionality | 3' to 5' (type α), 5' to 3' (type β) |

### 1.3.4 Ssl2's functional domains and motifs

**Table 3 Summary of published Ssl2 and its homolog structures**

| PDB number | Species | Protein | Year | Method | Full length | Assigned sequence in the structure | Description |
|---|---|---|---|---|---|---|---|
| 2FWR | *Archaeoglobus Fulgidis* | AfXPB | 2006 | X-RAY DIFFRACTION | 1-452 | N/A | Structure of *Archaeoglobus Fulgidis* XPB[64] |
| 2FZ4 | *Archaeoglobus Fulgidis* | AfXPB | 2006 | X-RAY DIFFRACTION | 1-452 | 24-229(206) | Crystal Structure of the N-terminal half of *Archaeoglobus Fulgidus* XPB[64] |
| 2FZL | *Archaeoglobus Fulgidis* | AfXPB | 2006 | X-RAY DIFFRACTION | 1-452 | 258-454(196) | Structure of C-terminal domain of *Archaeoglobus fulgidus* XPB[64] |
| 4ERN | *Homo sapiens* | XPB | 2013 | X-RAY DIFFRACTION | 1-782 | 502-730(289) | Crystal structure of the C-terminal domain of human XPB/ERCC-3 excision repair protein at 1.80 A[65] |
| 5OF4 | *Homo sapiens* | XPB | 2017 | ELECTRON MICROSCOPY | 1-782 | 266-728(503) | The cryo-EM structure of human TFIIH[66] |
| 5OQJ | *Saccharomyces cerevisiae* | Ssl2 | 2017 | ELECTRON MICROSCOPY | 1-843 | 363-425, 482-770 | Structure of yeast transcription pre-initiation complex with TFIIH[7] |
| 5OQM | *Saccharomyces cerevisiae* | Ssl2 | 2017 | ELECTRON MICROSCOPY | 1-843 | 363-425, 482-770 | Structure of yeast transcription pre-initiation complex with TFIIH and core mediator[7] |
| 6RO4 | *Homo sapiens* | XPB | 2019 | ELECTRON MICROSCOPY | 1-782 | 71-720(650) | Structure of the core TFIIH-XPA-DNA complex[67] |
| 6O9M | *Homo sapiens* | XPB | 2019 | ELECTRON MICROSCOPY | 1-782 | 30-726(697) | Structure of the human apo TFIIH[68] |
| 6O9L | *Homo sapiens* | XPB | 2019 | ELECTRON MICROSCOPY | 1-782 | 30-728(699) | Human holo-PIC in the closed state[68] |
| 6NMI | *Homo sapiens* | XPB | 2019 | ELECTRON MICROSCOPY | 1-782 | 34-203, 265-730 | Cryo-EM structure of the human TFIIH core complex[52] |

The structures of Ssl2 and its homologs' structures have been determined in different molecular contexts (summarized in **Table 3**)[7,52,64-68]. The first structure came from *Archaeoglobus fulgidus* XPB (AfXPB), an archaeal homolog of Ssl2, and revealed two primary sequence predicted RecA-like core domains at the center of the protein conserved in SF2 helicases, termed RecA-like domain 1 (RecA1) or helicase domain 1 (HD1), and RecA-like domain 2 (RecA2) or helicase domain 2 (HD2)[64] **(Figure 9)**. HD1 and HD2 structures of XPB and Ssl2 were solved later in human XPB and *S. cerevisiae* Ssl2, respectively, which in both organisms share high degree of structural similarity with AfXPB in *A. fulgidus*[7,52,64-68]. The AfXPB structure additionally uncovers a DNA damage recognition domain (DRD) at the N-terminal region that structurally resembles a DNA repair protein MutS and has been implicated to function in DNA damage recognition and also shares high sequence similarity in eukaryotic homologs[52,64]. Moreover, a RED motif, which is highly conserved across XPB proteins from archaeal to human, has been noted in the HD1 C-term region, which is composed of three continuous amino acids of Arginine, Glutamic acid and Aspartic acid (RED), and forms a loop structure the mutation of which dramatically reduces AfXPB's helicase activity **(Figure 11)**[64]. A flexible thumb-like motif (ThM) is identified within the AfXPB HD2 N-term region, which structurally resembles DNA polymerases with an implicated function in nucleic acid binding. A ThM motif is also found in XPB but shorter in length and deletion of the first half of XPB ThM is detrimental to XPB integration into TFIIH during DNA repair[69]. Compared to archaeal AfXPB, human XPB and *S. cerevisiae* Ssl2 contain longer N terminal domains (NTD) and C terminal extensions (C-ter) **(Figure 9)**. Biochemical and structural data of XPB and Ssl2 independently show that NTD is critical for XPB/Ssl2's integration to the TFIIH complex[7,70]. Two human disease-related mutations, F99S and T119P, are found in the NTD of XPB and these two mutations may reduce levels of active TFIIH by perturbing its

stability[52,71]. The Ssl2 C-terminal extension has been found to interact with TFIIH subunit Tfb5 in the Tfb2-Tfb5 dimerization region[7].

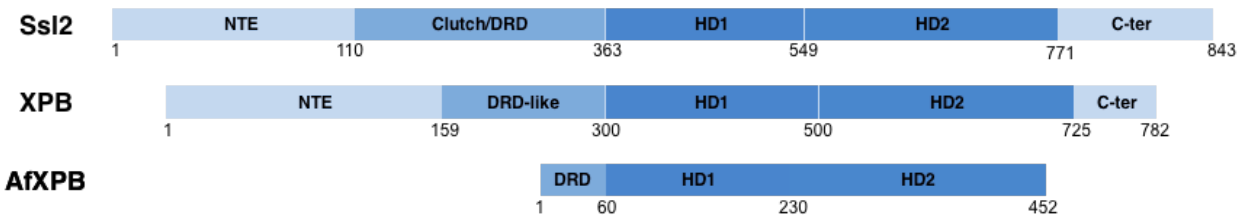



**Figure 9 Domains and motifs of Ssl2, XPB and AfXPB**

Adaped from Li Fan, S. Schilbach Chunli Yan *et. al*[7,64,68]. Ssl2, XPB and AfXPB all contain two helicase core domains, HD1 and HD2. AfXPB has a DRD motif at the N-ternimal, which also has been identified in Ssl2 and in XPB as the DRD-like motif. Ssl2 and XPB additionally possess an extended N-terminal domain (NTE) and a C-terminal extension (C-ter).

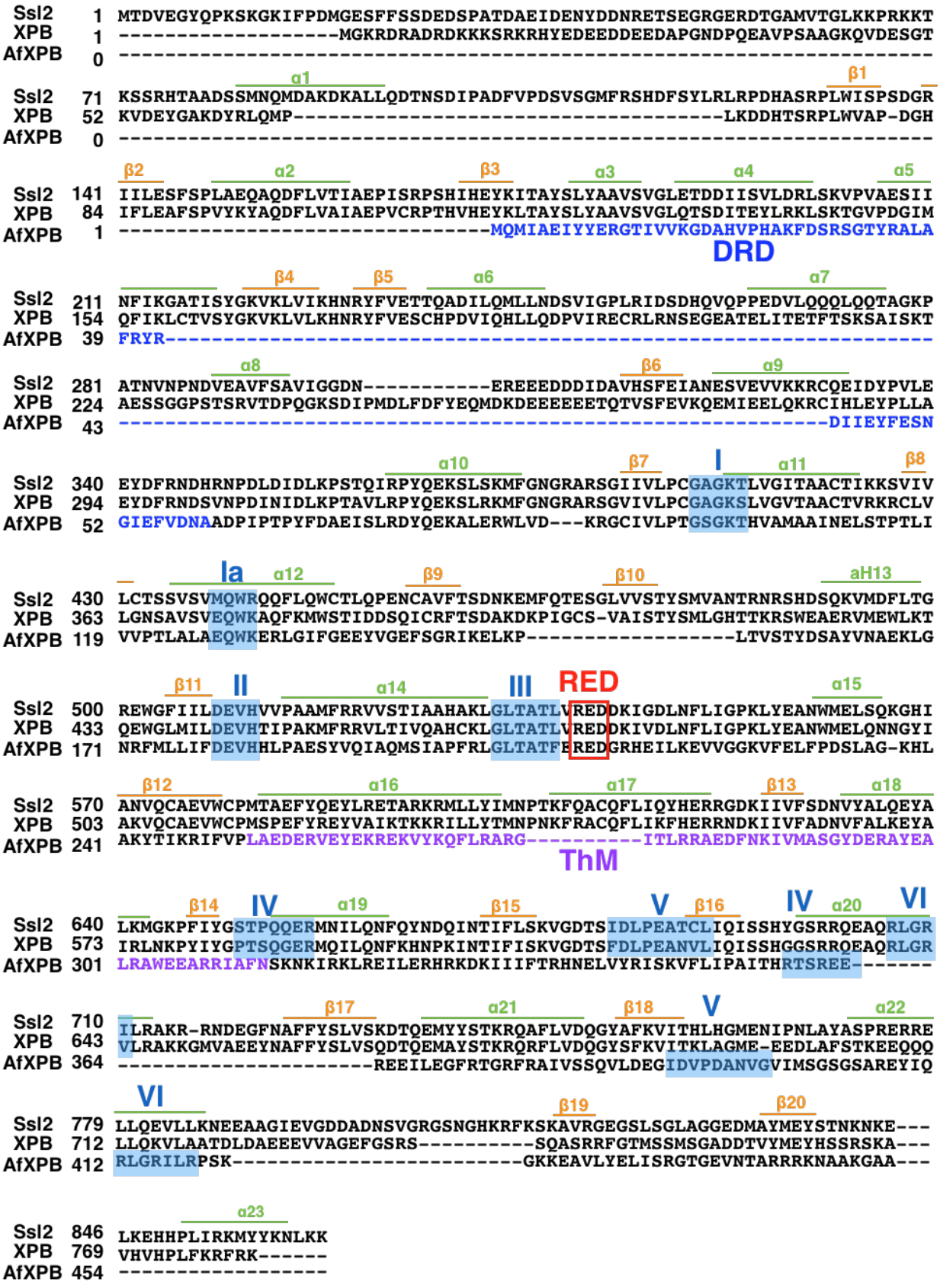

Ssl2 1 MTDVEGYQPKSKGKIFPDMGESFFSSDEDSPATDAEIDENYDDNRETSEGRGERDTGAMVTGLKKPRKKT
XPB 1 ----------MGKRDRADRDKKKSRKRHYEDEEDDEEDAPGNDPQEAVPSAAGKQVDESGT
AfXPB 0 ----------
α1 β1
Ssl2 71 KSSRHTAADSSMNQMDAKDKALLQDTNSDIPADFVPDSVSGMFRSHDFSYLRLRPDHASRPLWISPSDGR
XPB 52 KVDEYGAKDYRLQMP----------LKDDHTSRPLWVAP-DGH
AfXPB 0 ----------
β2 α2 β3 α3 α4 α5
Ssl2 141 IILESFSPLAEQAQDFLVTIAEPISRPSHIHEYKITAYSLYAAVSVGLETDDIISVLDRLSKVPVAESII
XPB 84 IFLEAFSPVYKYAQDFLVAIAEPVCRPTHVHEYKLTAYSLYAAVSVGLQTSDITEYLRKLSKTGVPDGIM
AfXPB 1 ----------MQMIAEIYYERGTIVVKGDAHVPHAKFDSRSGTYRALA
DRD
β4 β5 α6 α7
Ssl2 211 NFIKGATISYGKVKLVIKHNRYFVETTQADILQMLLNDSVIGPLRIDSDHQVQPPEDVLQQQLQQTAGKP
XPB 154 QFIKLCTVSYGKVKLVLKHNRYFVESCHPDVIQHLLQDPVIRECRLRNSEGEATELITETFTSKSAISKT
AfXPB 39 FRYR----------
α8 β6 α9
Ssl2 281 ATNVNPNDVEAVFSAVIGGDN----------EREEEDDDIDAVHSFEIANESVEVVKKRCQEIDYPVLE
XPB 224 AESSGGPSTSRVTDPQGKSDIPMDLFDFYEQMDKDEEEEEETQTVSFEVKQEMIEELQKRCIHLEYPLLA
AfXPB 43 ----------DIIEYFESN
α10 β7 I α11 β8
Ssl2 340 EYDFRNDHRNPDLDIDLKPSTQIRPYQEKSLSKMFGNGRARSGIIVLPCGAGKTLVGITAACTIKKSVIV
XPB 294 EYDFRNDSVNPDINIDLKPTAVLRPYQEKSLRKMFGNGRARSGVIVLPCGAGKSLVGVTAACTVRKRCLV
AfXPB 52 GIEFVDNAADPIPTPYFDAEISLRDYQEKALERWLVD---KRGCIVLPTGSGKTHVAMAAINELSTPTLI
Ia α12 β9 β10 aH13
Ssl2 430 LCTSSVSVMQWRQQFLQWCTLQPENCAVFTSDNKEMFQTESGLVVSTYSMVANTRNRSHDSQKVMDFLTG
XPB 363 LGNSAVSVEQWKAQFKMWSTIDDSQICRFTSDAKDKPIGCS-VAISTYSMLGHTTKRSWEAERVMEWLKT
AfXPB 119 VVPTLALAEQWKERLGIFGEEYVGEFSGRIKELKP----------LTVSTYDSAYVNAEKLG
β11 II α14 III RED α15
Ssl2 500 REWGFIILDEVHVVPAAMFRRVVSTIAAHAKLGLTATLVREDDKIGDLNFLIGPKLYEANWMELSQKGHI
XPB 433 QEWGLMILDEVHTIPAKMFRRVLTIVQAHCKLGLTATLVREDDKIVDLNFLIGPKLYEANWMELQNNGYI
AfXPB 171 NRFMLLIFDEVHHLPAESYVQIAQMSIAPFRLGLTATFEREDGRHEILKEVVGGKVFELFPDSLAG-KHL
β12 α16 α17 β13 α18
Ssl2 570 ANVQCAEVWCPMTAEFYQEYLRETARKRMLLYIMNPTKFQACQFLIQYHERRGDKIIVFSDNVYALQEYA
XPB 503 AKVQCAEVWCPMSPEFYREYVAIKTKKRILLYTMNPNKFRACQFLIKFHERRNDKIIVFADNVFALKEYA
AfXPB 241 AKYTIKRIFVPLAEDERVEYEKREKVYKQFLRARG----------ITLRRAEDFNKIVMASGYDERAYEA
ThM
β14 IV α19 β15 V β16 IV α20 VI
Ssl2 640 LKMGKPFIYGSTPQQERMNILQNFQYNDQINTIFLSKVGDTSIDLPEATCLIQISSHYGSRRQEAQRLGR
XPB 573 IRLNKPYIYGPTSQGERMQILQNFKHNPKINTIFISKVGDTSFDLPEANVLIQISSHGGSRRQEAQRLGR
AfXPB 301 LRAWEEARRIAFNSKNKIRKLREILERHRKDKIIIFTRHNELVYRISKVFLIPAITHRTSREE------
β17 α21 β18 V α22
Ssl2 710 ILRAKR-RNDEGFNAFFYSLVSKDTQEMYYSTKRQAFLVDQGYAFKVITHLHGMENIPNLAYASPRERRE
XPB 643 VLRAKKGMVAEEYNAFFYSLVSQDTQEMAYSTKRQRFLVDQGYSFKVITKLAGME-EEDLAFSTKEEQQQ
AfXPB 364 ----------REEILEGFRTGRFRAIVSSQVLDEGIDVPDANVGVIMSGSGSAREYIQ
VI
β19 β20
Ssl2 779 LLQEVLLKNEEAAGIEVGDDADNSVGRGSNGHKRFKSKAVRGEGSLSGLAGGEDMAYMEYSTNKNKE---
XPB 712 LLQKVLAATDLDAEEEVVAGEFGSRS----------SQASRRFGTMSSMSGADDTVYMEYHSSRSKA---
AfXPB 412 RLGRILRPSK----------GKKEAVLYELISRGTGEVNTARRRKNAAKGAA---
α23
Ssl2 846 LKEHHPLIRKMYYKNLKK
XPB 769 VHVHPLFKRFRK------
AfXPB 454 ----------

**Figure 10 Sequence alignment and motif annotation of Ssl2, XPB and AfXPB**

Protein sequences of Ssl2, XPB and AfXPB were aligned using CLUSTALW: https://www.genome.jp/tools-bin/clustalw. The secondary structure of Ssl2 was predicted using PSIPRED: http://bioinf.cs.ucl.ac.uk/psipred/, and labelled on top of the sequence, α stands for α-helix and β stands for β-strand. The DRD, RED and ThM motifs identified from Li Fan *et. al*[64] in AfXPB are highlighted. Conserved helicase motifs I-VI are annotated as indicated by Li Fan *et. al*[64].

### 1.3.5 Ssl2 architecture and its interactions with other PIC components

Ssl2/XPB assembles with other TFIIH subunits to form a horseshoe-like overall shape **(Figure 8)**[52]. The direct contact between Ssl2/XPB (yeast and human homologs, also applies to the following description) and the second largest TFIIH subunit Rad3/XPD is additionally bridged by MAT1/Tfb3 on one side of the complex. The other side of TFIIH structure contains other TFIIH components in a rough order of Rad3/XPD-Ssl1/p44-Tfb1/p62-Tfb4/p34-Tfb2/p52-Tfb5/p8-Ssl2/XPB[7,52,66]. The overall structures of Ssl2, XPB and AfXPB are very similar, containing two lobe-like modules, lobe 1 and lobe2, with lobe 1 centered at HD1 and lobe 2 centered at HD2 **(Figure 1, 9)**. Ssl2 and XPB additionally have NTD and DRD-like structures at the N-terminal ends, and the N-term extension at the C-terminal ends that have been found to interact with other PIC components. Domain-level interactions between TFIIH subunits are observed in the human TFIIH structure, for example, XPB's NTD contacts the p52 "clutch" domain through hydrophobic and charged interactions; the XPB DRD motif interacts with the XPD RecA2 domain and the MAT1 helical domain; Ssl2 RecA1 interacts with XPD RecA2 domain[52]. An interaction between Ssl2 NTD and Tfb5 is also observed in yeast PIC complex[7]. Additionally, as mentioned earlier, Ssl2 C-term is structurally visualized to interact with Tfb5's Tfb5-Tfb2 dimerization module in the yeast PIC. Moreover, Ssl2 lobe 2 contacts with TFIIE's C-terminal end in the yeast PIC[7].

Besides structurally visualized inter-subunit interactions of Ssl2 and other PIC components, amino-acid level interactions are revealed by BS3, SBAT and FeBABE-based protein crosslink experiments (summarized in **Table 4, Figure 10**)[7,45,48,72,73].

**Table 4 Summary of published protein crosslink data between Ssl2 and the other PIC components**

| Year | Description | | | |
|---|---|---|---|---|
| **2012** | **Summary of protein cleavages within the PIC using TFIIE-FeBABE probes[48]** | | | |
| | **Protein 1** | **Position 1** | **Protein 2** | **Position 2** |
| | Tfa1 | 56 | Ssl2 | 448 |
| | Tfa1 | 93 | Ssl2 | 448 |
| | Tfa1 | 137 | Ssl2 | 448 |
| | Tfa1 | 174 | Ssl2 | 450 |
| | Tfa1 | 174 | Ssl2 | 442 |
| | Tfa1 | 174 | Ssl2 | 553 |
| | Tfa1 | 174 | Ssl2 | 517 |
| | Tfa1 | 174 | Ssl2 | 612 |
| | Tfa1 | 174 | Ssl2 | 657 |
| | Tfa1 | 253 | Ssl2 | 444 |
| | Tfa1 | 263 | Ssl2 | 443 |
| | **Summary of protein crosslinks within the PIC using Ssl2-Bpa derivatives[48]** | | | |
| | **BPA insertion site** | **Identified crosslink targets** | | |
| | Ssl2 Q422 | TFIIB (strong) Tfa2 (weak) | | |
| | Ssl2 N434 | TFIIB (strong) Tfa2 (weak) | | |
| | Ssl2 S440 | TFIIB (strong) Tfa2 (weak) | | |
| | Ssl2 N442 | TFIIB (strong) Tfa2 (weak) | | |
| | Ssl2 G623 | Tfa2 | | |
| | Ssl2 Q634 | Tfa2 | | |
| | Ssl2 N638 | Tfa2 | | |
| **2013** | **Cross-links identified for the closed form of PIC[72]** | | | |
| | **Protein 1** | **Position 1** | **Protein 2** | **Position 2** |
| | Rpb1 | 1262 | Ssl2 | 839 |
| | Tfg2 | 148 | Ssl2 | 839 |
| | Rpb1 | 1246 | Ssl2 | 827 |
| | Tfg2 | 174 | Ssl2 | 839 |
| | Rpb1 | 1217 | Ssl2 | 827 |
| | Rpb1 | 1262 | Ssl2 | 835 |
| | **Cross-links identified for TFIIH[72]** | | | |
| | **Protein 1** | **Position 1** | **Protein 2** | **Position 2** |
| | Ssl2 | 228 | Tfb3 | 192 |
| | Ssl2 | 224 | Ssl1 | 52 |
| | Ssl2 | 794 | Tfb5 | 60 |
| | Ssl2 | 357 | Tfb1 | 238 |
| **2015** | **Identified crosslinks from yeast TFIIH[45]** | | | |

| | Protein 1 | Position 1 | Protein 2 | Position 2 |
|---|---|---|---|---|
| | Ssl2 | 734 | Tfb2 | 415 |
| | Tfb1 | 238 | Ssl2 | 357 |
| | Ssl2 | 357 | Tfb1 | 238 |
| | Tfb5 | 60 | Ssl2 | 721 |
| | Tfb6 | 289 | Ssl2 | 228 |
| | Ssl2 | 214 | Tfb6 | 162 |
| | Ssl2 | 357 | Tfb3 | 163 |
| | Ssl2 | 279 | Tfb3 | 226 |
| | Ssl2 | 228 | Tfb3 | 192 |
| | Ssl2 | 523 | Tfb3 | 219 |
| | Ssl2 | 64 | Tfb3 | 171 |
| | Rad3 | 30 | Ssl2 | 12 |
| | Rad3 | 499 | Ssl2 | 372 |
| | Ssl1 | 52 | Ssl2 | 228 |
| **2016** | **Cross-linked residue pairs identified from Med-PIC[73]** | | | |
| | **Protein 1** | **Position 1** | **Protein 2** | **Position 2** |
| | Rpb1 | 1262 | Ssl2 | 835 |
| | Rpb1 | 1262 | Ssl2 | 839 |
| | Rpb1 | 1217 | Ssl2 | 827 |
| | Rpb1 | 1246 | Ssl2 | 824 |
| | Rpb2 | 228 | Ssl2 | 824 |
| | Ssl2 | 694 | Med2 | 31 |
| | Ssl2 | 472 | Rpb5 | 94 |
| | Ssl2 | 228 | Ssl1 | 52 |
| | Ssl2 | 711 | Tfa1 | 301 |
| | Ssl2 | 10 | Tfb2 | 506 |
| | Ssl2 | 796 | Tfb2 | 490 |
| | Tfa2 | 277 | Ssl2 | 472 |
| | Tfa2 | 311 | Ssl2 | 796 |
| | Tfb5 | 6 | Ssl2 | 12 |
| **2017** | **EDC-crosslinks in PIC-Mediator core[7]** | | | |
| | **Protein 1** | **Position 1** | **Protein 2** | **Position 2** |
| | Ssl2 | 520 | Med7 | 1 |
| | Ssl2 | 351 | Ssl1 | 52 |
| | Tfb3 | 44 | Ssl2 | 69 |
| | Tfb5 | 6 | Ssl2 | 46 |

Study of the AfXPB structure suggests a large conformational change before and after DNA binding inferred from the structure comparison between AfXPB and a DNA bound helicase[64]. Indeed, a relative position change is observed between two XPB lobes by comparing

XPB within TFIIH and promoter bound PIC, however, there is no dramatic inter-domain rearrangements observed[7,66]. An obvious conformational change upon DNA binding occurs between XPB and XPD, during which the XPB-XPD interaction breaks and XPB moves together with its other interaction domains in p8 and p52, away from XPD **(Figure 11)**[66].

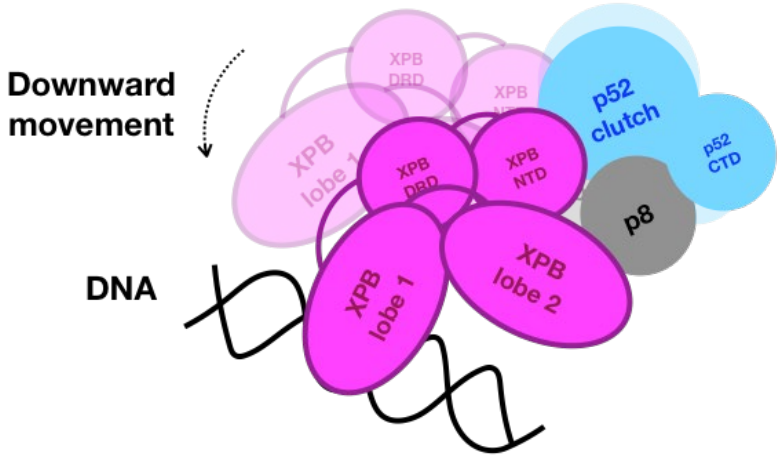



**Figure 11 Schematic illustration of the conformational change of XPB and its interaction domains between free TFIIH and PIC-bound TFIIH**

Adapted from Basil J Greber *et. al*[52]. The XPB/Ssl2-p8/Tfb5-p52/Tfb2 module moves downward upon DNA binding within the PIC complex.

## 1.4 ADDITIONAL FACTORS INVOLVED IN RNA POL II TSS SELECTION IN YEAST

Mutagenesis of the upstream duplicated Inr increases the efficiency of the downstream Inr and shifts the TSS distribution downstream[74,75]. The same effects of polarity on TSS distribution were also observed upon mutation of Pol II subunits Rpb1, Rpb2, Rpb7 and Rpb9, and GTFs TFIIB, TFIIF, TFIIH and the transcription co-activator Sub1. These results not only support a

unidirectional movement of the Pol II transcription machinery during initiation, but they also invoke new questions. What is the decisive factor for TSS selection? What factors are upstream to others in the pathway of TSS selection?

### 1.4.1 Pol II and TSS usage in yeast

Pol II, the enzyme that catalyzes RNA synthesis during transcription, is composed of 12 subunits, Rpb1-12, which are numbered from the largest to the smallest[34,76]. The largest subunit Rpb1 contains a highly conserved subdomain, the trigger loop (TL), which assists the catalytic center of Pol II for phosphodiester bond formation and substrate selection during transcription[75]. Studies from our lab and others have identified mutations in the Pol II TL that alter the measured transcription elongation rate *in vitro*. Pol II mutants with increased elongation rate are termed as gain of function (GOF) catalytic alleles, and mutants with decreased elongation rate are termed as loss of function (LOF) catalytic alleles. Our lab's study found that Pol II GOF alleles shifted TSS distribution upstream both at a model gene, *ADH1*, and genome-wide[75,77-79]. In contrast, Pol II LOF alleles shifted TSS distribution downstream at *ADH1* and genome-wide. Mutations in other subunits of Pol II have also been identified having effects on TSS selection, including *rpb2* G369S, which was discovered to suppress *sua7-3* (TFIIB R78C) downstream TSS shifts, and *rpb2* E437G/F442S, which shifts TSS usage upstream[77,80]; *rpb7* D166G, the first identified Rpb7 mutant that alters TSS usage[77]; and *ssu73-1* allele, which is a Rpb9 mutant that contains a nonsense mutation at codon 107 and encodes a truncated form of Rpb9 lacking the last 16 amino acids, rescuing *sua7-1* (TFIIB E62K) caused growth and TSS defects[81].

### 1.4.2 TFIIB and TSS selection

TFIIB is a key component of the PIC that functions in Pol II recruitment and transcription start site recognition during transcription initiation[34]. TFIIB contains four major domains, N-terminal B-ribbon, B-reader, B-linker, N-term B-core and C-term B-core, and the C-terminal tail[2]. The Pol II-TFIIB structure reveals that the entire TFIIB is positioned on Pol II[2,82]. The B-ribbon interacts with the Pol II "dock", B-reader contacts the Pol II "wall", B-linker interacts with the Pol II "clamp" coiled-coil, the N-term B-core associates with RNA exit tunnel, and the C-term B-core interacts with both TBP and DNA[83-85]. Based on published data, a model for TFIIB function in transcription initiation and the transition to elongation has been proposed[2,82]. In the early stage of initiation, TFIIB recruits Pol II to promoter DNA through interactions between the TFIIB B-ribbon and Pol II dock, and TFIIB B-core and Pol II wall. At the same time, DNA is positioned above the Pol II active center cleft. Binding of TFIIB to Pol II triggers structural changes of Pol II domains and the Pol II cleft is partially closed to position and help DNA opening. DNA melting then occurs at ~20 bp downstream of the TATA box with the assistance of the B-linker. The emerging DNA template strand slides into the template tunnel and the downstream double stranded DNA is loaded onto the cleft. TFIIB B-reader binds to the DNA template strand upstream to help with the recognition of the Inr sequence and TSS selection. B-reader residues R64 and D69 bind to the DNA template strand at positions -7/-8, with specific interaction between thymine at -8 and the B-reader proposed to be a critical recognition step for Pol II to find Inr and define TSS[82]. Multiple mutants have been found containing B-reader substitutions and cause downstream TSS shifts, including substitutions in E62, R64, R78, and V79. Additionally, B-reader mutations causing TSS shifts showed different sensitivities to changes of the Inr sequence. For example, R64A and -8A to -8T double mutant showed a great reduction in +1 TSS usage at *SNR14* promoter compared to

-8 mutant alone, suggesting that TFIIB mutations could affect Pol II TSS recognition through -8 position[20]. Another example showed that when an *ADH1* major TSS was mutagenized, V79L was able to increase the usage of this mutagenized TSS compared to the TSS mutated alone, indicating that TFIIB mutations could also confer altered Pol II specificity[86]. After Inr recognition and TSS selection, TFIIB has been proposed to promote the first phosphodiester bond formation by stimulating rearrangement of Pol II active-site residues and Pol II association with a catalytic $Mg^{2+}$ ion (Metal B). RNA synthesis results in the growth of the DNA-RNA hybrid within Pol II. When the RNA grows to 7 nucleotides, it separates from the DNA and clashes its interaction with three B-reader aspartate residues, D70, D74 and D75[82]. Once RNA grows to 12-13 nucleotides, it clashes with TFIIB, stimulating TFIIB release and elongation complex formation[82,87].

### 1.4.3 TFIIF and TSS selection

TFIIF in human is a heterodimer that is composed of two subunits RAP74 and RAP30, their homologous counterparts in yeast are Tfg1 and Tfg2[88]. Yeast TFIIF additionally contains a third subunit Tfg3, which is not essential and also serves as a component of several other remodeling complexes[34]. Human RAP74 and RAP30 both contain a dimerization domain at their N-termini. RAP74 has an acidic central region and followed by a winged-helix fold domain. The C-terminal region of RAP74 interacts with Pol II and TFIIB. RAP30 contains a central linker region that also interacts with Pol II, followed by a winged-helix domain. The C-terminal region of RAP30 possesses a DNA-binding activity, and the direct interaction between this domain and the promoter DNA is indicated to contribute to TFIIF's role in stabilizing the human PIC. Domain annotation of yeast Tfg1 and Tfg2 is originated from the understanding of human TFIIF subunits, based on the sequence conservation between the two homologous proteins. The most recent yeast

PIC structure containing TFIIF reveals that TFIIF locates inside the upstream of the PIC and makes direct contacts with DNA, Pol II, TFIIB and TFIIE[89]. TFIIF adopts an intricate fold that undergoes folding transitions with other GTFs to stabilize the opening complex. Mutations have been found in Tfg1 and Tfg2 that alter TSS usage, including two *TFG1* mutations, G363D (allele *ssu71-1*) and G363R (allele *ssu71-2*), which suppress *sua7-1*-encoded TFIIB E62K-caused growth and TSS defects[90,91]. Single mutation in Tfg2 L59K, shifts TSS upstream at *ADH1*[92]. Additionally, deletions of small pieces in the Tfg2 insertion motif (*tfg2Δ146-180*) or linker motif (*tfg2Δ233-248*) also shift TSS upstream[78].

### 1.4.4 Sub1 and TSS selection

Sub1 was originally identified as a high-copy suppressor of TFIIB R78H, a substitution contained within the B-reader motif conferring both cold sensitive growth and downstream TSS-shifting phenotypes[80,93]. Yeast Sub1 and its human homolog PC4 were both found to stimulate *in vitro* transcription and were defined as transcription co-activators[94,95]. In addition, Sub1 binds to promoters and was proposed to interact and stabilize the non-template DNA strand during initition[96]. Synthetic lethality between deletion of *SUB1* and two mutant alleles of TFIIB, E62G and R78H was also found[93]. Our lab's previous study also found widespread lethality *between sub1*Δ and two classes of downstream shifting initiation mutants: Pol II catalytic activity LOF alleles and the TFIIB *sua7-1* (E62K) allele[78]. Additionally, *sub1Δ* showed epistatic interactions with Pol II catalytic GOF alleles, distinct from the broad suppressive and additive interactions between Pol II, TFIIB and TFIIF alleles[77,78].

## 1.5 SUMMARY

As a subunit of TFIIH, Ssl2 has been demonstrated to use its translocase activity to open double stranded DNA during transcription initiation and is implicated in promoter scanning. Additionally, single amino acid substitutions have been found to affect TSS usage at a few of model genes. However, how Ssl2's translocase activity coordinates with Pol II's catalytic activity of nucleotide incorporation in determining TSS distribution is unclear. In addition, how does Ssl2 affect TSS distribution at all promoters has not been tested? Lastly, previous studies demonstrated that *S. cerevisiae* Pol II uses a scanning mechanism for transcription initiation and TSS selection, whether or not scanning is also a mechanism used by Pol II in other organisms is not studied. To answer the above questions, we used genetic tools to test the TSS phenotypes of several existing *ssl2* mutants and also screened for new *ssl2* mutants that are defective in TSS usage (Chapter II). We characterized these *ssl2* mutants and proposed a model to explain the possible mechanism for how *ssl2* coordinates with Pol II in determining TSS distribution (Chapter III). We used next generation sequencing (NGS) approach to evaluate Ss2's effect on TSS usage on promoters genome-wide (Chapter IV). We also initiated examination of the conservation of the scanning mechanism in *D. melanogaster* and *S. pombe* Pol II using NGS approaches (Appendix A).

# 2.0 SCREENING AND CHARACTERIZATION OF *SSL2* MUTANTS

## 2.1 INTRODUCTION

Studies have shown that TFIIH acts in open complex formation and is proposed to function TSS scanning in *S. cerevisiae* by using the DNA translocase activity of its Ssl2 subunit. *ssl2* alleles with mutations in the helicase domain were found to cause transcription-related growth defects while also altering TSS usage at model genes of *ADH1* and *CYC1*[54,56]. For example, *ssl2-508* (H508R), *ssl2-rtt* (E556K), *ssl2-DEAD* (V490A/H491D) and *SSL2-1* (W427L) display Gal$^-$ phenotypes, which are often related to defects in transcription initiation[56]. In addition, *ssl2-508* suppressed a TFIIB E62K mutant's growth and TSS defects. This TFIIB allele by itself shifts TSSs downstream, while *ssl2-508* suppresses this defect by shifting TSS usage upstream[56]. *ssl2-DEAD* similarly decreases TSS usage at the downstream positions of *ADH1*, measured by comparing usage of two major TSS 37 bp and 27 bp upstream from the *ADH1* start codon. Moreover, *rad25-ts24* (V552I/E556K) was found defective in RNA synthesis[54].

In addition to *SSL2* mutations that are known to cause transcriptional defects in yeast, mutations have also been found in human *XPB* that cause inherited autosomal recessive disorders xeroderma pigmentosum (XP), trichothiodystrophy (TTD), and Cockayne syndrome (CS)[97,98]. Disease-relevant mutations at *XPB* exons are summarized, and the analogous mutations in yeast are annotated here **(Table 5)**.

**Table 5 Disease related mutations in Human *XPB* and their corresponding yeast mutations**

| OMIM[a] | Mutation in *S. cerevisiae* | Mutation in *H. sapiens* |
|---|---|---|
| 0002 | F156S | F99S |
| 0003 | T176P | T119P |
| 0004 | K472* | R425* |
| 0005 | F316* | F270* |
| 0006 | D521* | D474* |
| 0007 | Q592* | Q545* |
| 0008 | S219* | S162* |
| N/A | Y750* | 4 bp insertion starting at amino acid 740 |

a. These disease mutations found in human *XPB* are also summarized at http://www.omim.org/entry/133510?search=ercc3&highlight=ercc3 except for Y750*, which in yeast is a mutant that contains a C-terminal truncated Ssl2 protein and originally constructed to mimic the truncated form of XPB identified from a XP patient[99,100]. Y750* in yeast was found to confer a putative transcription initiation defect related to a $GAL^-$ phenotype[56].

To further understand Ssl2's function in transcription initiation and TSS selection, we first characterized existing *ssl2* mutants in yeast and those analogous to mutants identified from other organisms. We then did a genetic screening to look for additional *ssl2* mutants that were defective in transcription initiation and TSS selection. Our lab developed genetic reporters that can detect transcriptional defects through conditional phenotypes, allowing us to classify mutants into functional groups. We also analyze TSS distributions at the model gene *ADH1* by quantification, which allows us to sensitively detect changes in TSS usage[75,78,101,102].

## 2.2 MATERIALS AND METHODS

### 2.2.1 Yeast strains

Yeast strains used in this chapter for study of the existing *ssl2* mutants are listed in **Table 6**.

**Table 6 Yeast strains used in studies of the existing *ssl2* mutants in Chapter II**

| Strain number | Plamids | Plasmid number | Genotype |
|---|---|---|---|
| CKY263 | | | *MATalpha ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56* |
| CKY860 | | | *MATa leu2Δ0 or Δ1 ura3-52 his3Δ200 trp1Δ63 met15Δ0 lys2-128∂ RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY1471 | pRSII316 *SSL2* | pCK1478 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY1475 | pRSII316 *SSL2* | pCK1478 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY1595 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY1596 | pRS315 *ssl2* E556K | pCK1501 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY1597 | pRS315 *ssl2* V490A/H491D | pCK1502 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY1598 | pRS315 *ssl2* W427L | pCK1503 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY1600 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY1601 | pRS315 *ssl2* E556K | pCK1501 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY1602 | pRS315 *ssl2* V490A/H491D | pCK1502 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY1603 | pRS315 *ssl2* W427L | pCK1503 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY1652 | pRS315 *ssl2* H508R | pCK1518 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY1653 | pRS315 *ssl2* V552I/E556K | pCK1519 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY1654 | pRS315 *ssl2* H508R | pCK1518 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY1655 | pRS315 *ssl2* V552I/E556K | pCK1519 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2410 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2411 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2412 | pRS315 *ssl2* F156S | pCK1820 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2413 | pRS315 *ssl2* F156S | pCK1820 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2414 | pRS315 *ssl2* Y750* | pCK1822 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2415 | pRS315 *ssl2* Y750* | pCK1822 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2416 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2417 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2418 | pRS315 *ssl2* F156S | pCK1820 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2419 | pRS315 *ssl2* F156S | pCK1820 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2420 | pRS315 *ssl2* Y750* | pCK1822 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2421 | pRS315 *ssl2* Y750* | pCK1822 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2473 | pRS315 *ssl2* T176P | pCK1996 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2474 | pRS315 *ssl2* T176P | pCK1996 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2489 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2490 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2491 | pRS315 *ssl2* T176P | pCK1996 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2492 | pRS315 *ssl2* T176P | pCK1996 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2507 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2508 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |

*Multiple WT strains are showed, each duplicate was independently transformed as control in different transformation batches.

Yeast strains used for *ssl2* mutant screening and related experiments are listed in **Table 7**.

**Table 7 Yeast strains used for *ssl2* mutant sreening and related experiments in Chapter III**

| Strain number | Plamids | | Genotype |
|---|---|---|---|
| CKY2238 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2239 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2240 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2241 | pRS315 *SSL2* | pCK1480 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2242 | pRS315 *ssl2* I170T | pCK1826 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2243 | pRS315 *ssl2* I170T | pCK1826 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2244 | pRS315 *ssl2* I170T | pCK1826 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2245 | pRS315 *ssl2* I170T | pCK1826 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2246 | pRS315 *ssl2* L180S | pCK1827 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2247 | pRS315 *ssl2* L180S | pCK1827 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2248 | pRS315 *ssl2* L180S | pCK1827 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2249 | pRS315 *ssl2* L180S | pCK1827 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2250 | pRS315 *ssl2* S185Y | pCK1828 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2251 | pRS315 *ssl2* S185Y | pCK1828 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2252 | pRS315 *ssl2* S185Y | pCK1828 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2253 | pRS315 *ssl2* S185Y | pCK1828 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2254 | pRS315 *ssl2* S201Y | pCK1829 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2255 | pRS315 *ssl2* S201Y | pCK1829 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2256 | pRS315 *ssl2* S201Y | pCK1829 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2257 | pRS315 *ssl2* S201Y | pCK1829 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2258 | pRS315 *ssl2* I218T | pCK1830 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2259 | pRS315 *ssl2* I218T | pCK1830 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2260 | pRS315 *ssl2* I218T | pCK1830 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2261 | pRS315 *ssl2* I218T | pCK1830 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2262 | pRS315 *ssl2* L225P | pCK1831 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2263 | pRS315 *ssl2* L225P | pCK1831 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2264 | pRS315 *ssl2* L225P | pCK1831 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2265 | pRS315 *ssl2* L225P | pCK1831 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2266 | pRS315 *ssl2* V226D | pCK1832 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2267 | pRS315 *ssl2* V226D | pCK1832 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2268 | pRS315 *ssl2* V226D | pCK1832 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2269 | pRS315 *ssl2* V226D | pCK1832 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2270 | pRS315 *ssl2* N230S | pCK1833 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2271 | pRS315 *ssl2* N230S | pCK1833 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2272 | pRS315 *ssl2* N230S | pCK1833 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2273 | pRS315 *ssl2* N230S | pCK1833 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2274 | pRS315 *ssl2* N230I | pCK1834 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2275 | pRS315 *ssl2* N230I | pCK1834 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2276 | pRS315 *ssl2* N230I | pCK1834 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2277 | pRS315 *ssl2* N230I | pCK1834 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2278 | pRS315 *ssl2* N230D | pCK1835 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2279 | pRS315 *ssl2* N230D | pCK1835 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2280 | pRS315 *ssl2* N230D | pCK1835 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2281 | pRS315 *ssl2* N230D | pCK1835 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2282 | pRS315 *ssl2* Y232D | pCK1836 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2283 | pRS315 *ssl2* Y232D | pCK1836 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2284 | pRS315 *ssl2* Y232D | pCK1836 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2285 | pRS315 *ssl2* Y232D | pCK1836 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2286 | pRS315 *ssl2* L246P | pCK1837 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2287 | pRS315 *ssl2* L246P | pCK1837 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2288 | pRS315 *ssl2* L246P | pCK1837 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2289 | pRS315 *ssl2* L246P | pCK1837 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2290 | pRS315 *ssl2* V268A | pCK1838 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2291 | pRS315 *ssl2* V268A | pCK1838 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2292 | pRS315 *ssl2* V268A | pCK1838 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2293 | pRS315 *ssl2* V268A | pCK1838 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2294 | pRS315 *ssl2* N320Y | pCK1839 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2295 | pRS315 *ssl2* N320Y | pCK1839 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2296 | pRS315 *ssl2* N320Y | pCK1839 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2297 | pRS315 *ssl2* N320Y | pCK1839 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2298 | pRS315 *ssl2* K328N | pCK1840 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2299 | pRS315 *ssl2* K328N | pCK1840 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2300 | pRS315 *ssl2* K328N | pCK1840 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2301 | pRS315 *ssl2* K328N | pCK1840 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2302 | pRS315 *ssl2* E340G | pCK1841 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2303 | pRS315 *ssl2* E340G | pCK1841 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2304 | pRS315 *ssl2* E340G | pCK1841 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2305 | pRS315 *ssl2* E340G | pCK1841 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2306 | pRS315 *ssl2* N345T | pCK1842 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2307 | pRS315 *ssl2* N345T | pCK1842 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2308 | pRS315 *ssl2* N345T | pCK1842 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2309 | pRS315 *ssl2* N345T | pCK1842 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2310 | pRS315 *ssl2* D346G | pCK1843 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2311 | pRS315 *ssl2* D346G | pCK1843 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2312 | pRS315 *ssl2* D346G | pCK1843 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2313 | pRS315 *ssl2* D346G | pCK1843 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2314 | pRS315 *ssl2* N349D | pCK1844 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2315 | pRS315 *ssl2* N349D | pCK1844 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2316 | pRS315 *ssl2* N349D | pCK1844 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2317 | pRS315 *ssl2* N349D | pCK1844 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2318 | pRS315 *ssl2* K357R | pCK1845 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2319 | pRS315 *ssl2* K357R | pCK1845 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2320 | pRS315 *ssl2* K357R | pCK1845 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2321 | pRS315 *ssl2* K357R | pCK1845 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2322 | pRS315 *ssl2* K357E | pCK1846 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2323 | pRS315 *ssl2* K357E | pCK1846 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2324 | pRS315 *ssl2* K357E | pCK1846 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2325 | pRS315 *ssl2* K357E | pCK1846 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2326 | pRS315 *ssl2* K372E | pCK1847 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2327 | pRS315 *ssl2* K372E | pCK1847 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2328 | pRS315 *ssl2* K372E | pCK1847 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2329 | pRS315 *ssl2* K372E | pCK1847 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2330 | pRS315 *ssl2* G382V | pCK1848 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2331 | pRS315 *ssl2* G382V | pCK1848 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2332 | pRS315 *ssl2* G382V | pCK1848 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2333 | pRS315 *ssl2* G382V | pCK1848 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2334 | pRS315 *ssl2* I383T | pCK1849 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2335 | pRS315 *ssl2* I383T | pCK1849 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2336 | pRS315 *ssl2* I383T | pCK1849 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2337 | pRS315 *ssl2* I383T | pCK1849 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2338 | pRS315 *ssl2* T402I | pCK1850 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2339 | pRS315 *ssl2* T402I | pCK1850 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2340 | pRS315 *ssl2* T402I | pCK1850 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2341 | pRS315 *ssl2* T402I | pCK1850 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2342 | pRS315 *ssl2* K443E | pCK1851 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2343 | pRS315 *ssl2* K443E | pCK1851 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2344 | pRS315 *ssl2* K443E | pCK1851 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2345 | pRS315 *ssl2* K443E | pCK1851 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2346 | pRS315 *ssl2* V473D | pCK1852 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2347 | pRS315 *ssl2* V473D | pCK1852 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2348 | pRS315 *ssl2* V473D | pCK1852 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2349 | pRS315 *ssl2* V473D | pCK1852 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2350 | pRS315 *ssl2* F498L | pCK1853 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2351 | pRS315 *ssl2* F498L | pCK1853 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2352 | pRS315 *ssl2* F498L | pCK1853 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2353 | pRS315 *ssl2* F498L | pCK1853 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2354 | pRS315 *ssl2* D522V | pCK1854 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2355 | pRS315 *ssl2* D522V | pCK1854 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2356 | pRS315 *ssl2* D522V | pCK1854 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2357 | pRS315 *ssl2* D522V | pCK1854 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2358 | pRS315 *ssl2* K523N | pCK1855 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2359 | pRS315 *ssl2* K523N | pCK1855 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2360 | pRS315 *ssl2* K523N | pCK1855 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2361 | pRS315 *ssl2* K523N | pCK1855 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2362 | pRS315 *ssl2* N528I | pCK1856 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2363 | pRS315 *ssl2* N528I | pCK1856 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2364 | pRS315 *ssl2* N528I | pCK1856 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2365 | pRS315 *ssl2* N528I | pCK1856 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2366 | pRS315 *ssl2* F529L | pCK1857 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2367 | pRS315 *ssl2* F529L | pCK1857 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2368 | pRS315 *ssl2* F529L | pCK1857 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2369 | pRS315 *ssl2* F529L | pCK1857 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2370 | pRS315 *ssl2* E537G | pCK1858 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2371 | pRS315 *ssl2* E537G | pCK1858 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2372 | pRS315 *ssl2* E537G | pCK1858 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2373 | pRS315 *ssl2* E537G | pCK1858 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2374 | pRS315 *ssl2* W558R | pCK1859 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2375 | pRS315 *ssl2* W558R | pCK1859 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2376 | pRS315 *ssl2* W558R | pCK1859 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2377 | pRS315 *ssl2* W558R | pCK1859 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2378 | pRS315 *ssl2* D610A | pCK1860 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2379 | pRS315 *ssl2* D610A | pCK1860 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2380 | pRS315 *ssl2* D610A | pCK1860 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2381 | pRS315 *ssl2* D610A | pCK1860 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2382 | pRS315 *ssl2* R636C | pCK1861 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2383 | pRS315 *ssl2* R636C | pCK1861 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2384 | pRS315 *ssl2* R636C | pCK1861 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2385 | pRS315 *ssl2* R636C | pCK1861 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2386 | pRS315 *ssl2* N642D | pCK1862 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2387 | pRS315 *ssl2* N642D | pCK1862 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2388 | pRS315 *ssl2* N642D | pCK1862 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |

| | | | |
|---|---|---|---|
| CKY2389 | pRS315 *ssl2* N642D | pCK1862 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2390 | pRS315 *ssl2* L664S | pCK1863 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2391 | pRS315 *ssl2* L664S | pCK1863 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2392 | pRS315 *ssl2* L664S | pCK1863 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2393 | pRS315 *ssl2* L664S | pCK1863 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2394 | pRS315 *ssl2* Q672R | pCK1864 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2395 | pRS315 *ssl2* Q672R | pCK1864 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2396 | pRS315 *ssl2* Q672R | pCK1864 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2397 | pRS315 *ssl2* Q672R | pCK1864 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2398 | pRS315 *ssl2* E715G | pCK1865 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2399 | pRS315 *ssl2* E715G | pCK1865 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2400 | pRS315 *ssl2* E715G | pCK1865 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2401 | pRS315 *ssl2* E715G | pCK1865 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2402 | pRS315 *ssl2* S719P | pCK1866 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2403 | pRS315 *ssl2* S719P | pCK1866 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2404 | pRS315 *ssl2* S719P | pCK1866 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2405 | pRS315 *ssl2* S719P | pCK1866 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2406 | pRS315 *ssl2* L832* | pCK1867 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2407 | pRS315 *ssl2* L832* | pCK1867 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2Δ::hph* |
| CKY2408 | pRS315 *ssl2* L832* | pCK1867 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |
| CKY2409 | pRS315 *ssl2* L832* | pCK1867 | *MATa ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 RPB3::TAP::KlacTRP1 ssl2Δ::hph* |

### 2.2.2 Yeast media

Yeast media used in this study were made as previously described[75,78,102,103]. A brief summary, YP media is made of yeast extract (1% w/v; BD) and peptone (2% w/v; BD). Solid YP media was additionally supplemented with bacto agar (2% w/v; BD), adenine (0.15 mM, Sigma-Aldrich) and tryptophan (0.4 mM, Sigma-Aldrich). YPD media uses YP media components supplemented with dextrose (2% w/v, VWR). YPRaf media uses YP media components supplemented with raffinose (2% w/v, Amresco) and antimycin A (1 mg/ml; Sigma-Aldrich). YPRafGal media uses YP media components and supplemented with raffinose (2% w/v), galactose (1% w/v; Amresco) and antimycin A (1 mg/ml; Sigma-Aldrich). Minimal media (SC-) was made with a slightly modified "Hopkins mix" (0.2% most amino acids w/v), and supplemented with Yeast Nitrogen Base containing ammonium sulfate (without amino acids), bacto agar (2% w/v; BD) and dextrose (2% w/v, VWR). The original "Hopkins mix" and the slight modification were described[75,103]. The components of a batch of "Modified Hopkins mix" are Adenine Hemisulfate (2g), L-Alanine (2g), p-aminobenzoic acid (PABA) (0.2g), L-Arginine, L-Asparagine (2g), L-Aspartic acid (2g), L-Cysteine (2g), L-Glutamic Acid (2g), L-Glutamine (2g), Glycine (2g), L-Histidine (2g), Myo-Inositol (0.1g), L-Isoleucine (2g), L-Leucine (4g), L-Lysine (2g), L-Methionine (2g), L- Phenylalanine (2g), L-Proline (2g), L-Serine (2g), L-Threonine (2g), L-Tyrosine (2g), L-Tryptophan (2g), Uracil (2g), L-Valine (2g). All amino acids are from Sigma-Aldrich. SC-Leu+5FOA media is minimal media of SC-Leu supplemented with 5-fluoroorotic acid (5-FOA), (1 mg/ml, Gold Biotechnology). SC-Leu+MPA media is minimal media of SC-Leu supplemented with mycophenolic acid (20 ug/ml, Sigma-Aldrich). SC-His+3AT media is minimal media of SC-His supplemented with 3-aminotriazole (3-AT, 0.5mM, Sigma-Aldrich).

### 2.2.3 Plasmid shuffling and patch assay

To obtain cells with mutant *ssl2* plasmids as the sole source of *SSL2* gene in a haploid yeast strain, we performed plasmid shuffling as follows. A yeast strain was constructed where the genomic copy of *SSL2* was deleted while its essential functions were provided by *SSL2* cloned into a pRS316 *URA3* plasmid introduced into the strain. These cells were then transformed with a pRS315 *LEU2* plasmid, with either WT *SSL2* (used as a control) or mutagenized *ssl2* (*ssl2**). After transformation, yeast cells contain two copies of *SSL2* genes: WT *SSL2* on pRS316 *URA3* and WT *SSL2*/mutagenized *ssl2** on pRS315 *LEU2*. A patch assay was then performed to select yeast cells that had lost pRS316 *URA3* and retained pRS315 *LEU2*, so the phenotypes of *SSL2*/*ssl2** on the pRS315 *LEU2* plasmid would be apparent.

### 2.2.4 Plate phenotyping and growth heatmap

Yeast phenotyping assays were performed by spotting 10-fold serial dilutions of saturated YPD-liquid yeast cultures on various solid media, as previously described[75]. Yeast growth on specific media was recorded by taking pictures every 24 hours after an initial 16h of growth, from day 2 (40h) to day 7 for all media except for YPRaf/Gal (pictures to day 9). Growth phenotypes on specific media were scored on days when WT yeast reached mature colony sizes, as follows: YPD on day 2 (40h after spotting); SC-Leu, SC-His and SC-Trp on day 3 (64h); YPRaf on day 4 (88h); SC-Lys and SC-Leu+MPA on day 5 (112h); and YPRaf/Gal on day 7 (160h). To show the strength and distribution of mutants on the two-dimensional structure of Ssl2, growth phenotypes are converted to a numerical score, using the scale 1-5 to indicate the level of growth, where 1 indicates no growth and 5 indicates full growth. The level of growth is positively correlated with

the strength of phenotypes for SC-His, SC-Lys and YPRaf/Gal medium, so the "growth score" is directly used as "phenotyping score" for making a heat map. For other media, the level of growth is negatively correlated with the strength of growth, thus growth score 1-5 is inversely converted to the phenotyping strength score 5-1, with 5 growth score converted into phenotyping score 1 to indicate no phenotypes, with 4 growth score converted into phenotyping score 2 to show a weak phenotype, and so on. The heatmap uses light to dark color showing weak (phenotyping score 2) to strong phenotypes (phenotyping score 5), no phenotype (phenotyping score 0) is not shown.

### 2.2.5 Primer extension

To detect putative usage of TSSs in yeast, a primer extension (PE) assay was performed as previously described, with slight modification described in[75,104]. Briefly, 30μg of total RNA isolated from yeast cells was used for each PE reaction. A gene-specific oligo that anneals to the downstream site of the target gene promoter (usually 40-100 downstream of major TSS) was end-labeled with gamma-P32 ATP and T4 PNK, and annealed to the RNA. Reverse-transcription was then performed by adding M-MLV reverse-transcriptase (Fermentas) and RNase Inhibitor (Fermentas). RNase A was added to remove RNA template after reverse-transcription. Products were detected by running a 8% acrylamide gel made with 19:1 acrylamide:bisacrylamide (Bio-Rad), 1×TBE, and 7M urea. PE gels were visualized by phosphorimaging (GE Healthcare or Bio-Rad) and quantified by ImageQuant 5.1 (GE) or Image Lab software (Bio-Rad).

## 2.3 RESULTS

### 2.3.1 Yeast reporter alleles allow detection of transcriptional defects

Our lab has developed and used genetic reporters to detect transcription-dependent phenotypes and predict associated transcriptional defects. These reporters are *IMD2*, *imd2Δ::HIS3*, *gal10Δ56*, and *lys2-128∂* **(Figure 12)**. **(1)** Specifically, some mutants with transcription defects exhibit drug mycophenolic acid (MPA) sensitivity phenotypes ($MPA^S$), that correlate with altered TSS usage at *IMD2*. Under normal conditions, WT yeast use TATA-proximal GTP-initiated (first RNA nucleotide) TSSs at the *IMD2* promoter to initiate transcription, which produce non-functional transcripts due to presence of a downstream terminator preceding the *IMD2* open reading frame **(Figure 12A)**[105]. MPA promotes WT yeast to shift TSS usage from upstream GTP-initiated TSSs to a downstream ATP-initiated TSS, which generates a functional *IMD2* transcript. The Imd2 protein translated from this transcript is required for yeast to tolerate MPA. Thus, WT yeast are viable both in the absence and presence of drug MPA. However, transcription mutants that constitutively use the upstream *IMD2* TSSs or are unable to shift TSS usage to the downstream "A" site become sensitive to MPA. For example, our lab previously identified a set of Pol II mutants with increased elongation rate, termed gain of function (GOF) mutants[75,77]. In the presence of MPA, these Pol II GOF mutants only are able to shift TSS usage at *IMD2* to sites between TATA-proximal "G" TSSs and the downstream "A" TSS. Initiation at these intermediate sites also generates non-functional *IMD2* transcripts that will be rapidly degraded by the nuclear exosome, rendering Pol II GOF alleles MPA-sensitive[102]. Pol II GOF mutants with $MPA^S$ phenotypes were also found to shift TSS distributions upstream at *ADH1* and genome-wide, as detected by primer extension (PE) and genome-wide sequencing, respectively[75,79]. Strikingly, this correlation between

the MPA$^S$ phenotype and upstream TSS shifting was consistent across all our tested TSS mutants. **(2)** Conversely, we found previously that mutants shifting TSSs downstream constitutively express *IMD2* in the absence of using MPA as inducer[75]. To detect mutants of this class, we created a reporter *imd2Δ::HIS3*, where the *IMD2* ORF is replaced by the *HIS3* ORF; thus, *HIS3* mRNA will be produced to support yeast growth on SC-His medium when the downstream *IMD2* TSS is used constitutively **(Figure 12B)**[102]. Expression of *HIS3* will confer a His$^+$ phenotype and is indicative of TSS downstream shifting. Indeed, Pol II mutants with decreased elongation rate and termed loss of function (LOF) catalytic mutants display His$^+$ phenotypes and shift TSS distribution downstream, both at *ADH1* and genome-wide[75,102]. **(3)** Mutants that suppress the reduction in *GAL7* expression caused by deletion of the *GAL10* polyadenylation signal (*gal10Δ56* reporter) suppress galactose toxicity caused by *gal10Δ56* (Galactose resistance=Gal$^R$). Deletion of the major *GAL10* polyadenylation signal interferes with *GAL10* 3'-end formation causing transcription read-through and interference with initiation from the downstream *GAL7* gene **(Figure 12C)**. Reduction of *GAL7* expression in the presence of galactose results in the buildup of a toxic metabolite, which would normally be prevented by Gal7 enzymatic activity, causing a galactose toxicity. Transcription mutants that either increase *GAL10* 3'-end formation, termination downstream of *GAL10*, or *GAL7* initiation directly or indirectly are predicted to suppress galactose toxicity and exhibit Gal$^R$ phenotypes, which were often seen in our lab's previous Pol II mutant screening[75]. **(4)** Mutants that suppress a *Ty* element insertion in *LYS2* (*lys2-128∂*) confer a Spt$^-$ phenotype. Insertion of a Ty element at the 5' end of *LYS2* blocks Pol II read-through and results in a short non-functional transcript and lysine auxotrophy **(Figure 12D)**[106]. Certain mutants that allow expression of *LYS2* from a cryptic promoter thereby supporting yeast growth on SC-Lys medium (the Spt$^-$ phenotype), and has been observed in a subset of the strongest Pol II GOF mutants[75].

These transcriptional phenotypes together with heat-sensitive (Tsm$^-$) and cold-sensitive (Csm$^-$) phenotypes at restrictive temperatures, allow us to potentially classify mutants into different functional groups.

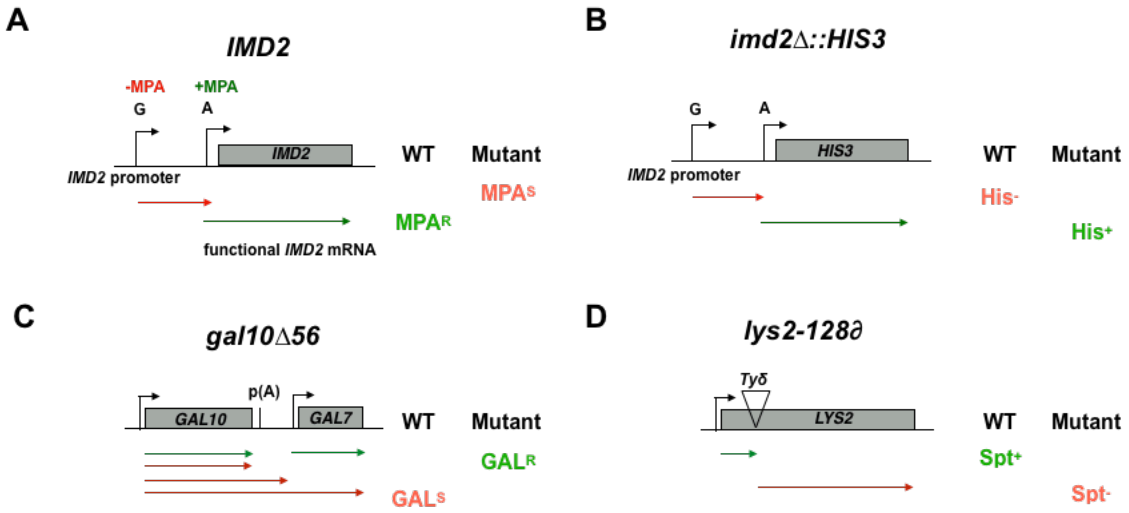



**Figure 12 Yeast genetic reporters allow detection of transcriptional defects**

**(A)** Schematic illustration of *IMD2* promoter and its two conditional transcripts. In the absence MPA, transcription initiates at upstream "G" sites and produces non-functional *IMD2* transcripts. In the presence of MPA, transcription initiates at the downstream "A" site and generates a functional transcript required for yeast to tolerate MPA. Inability to shift to the downstream TSS in the presence of MPA leads to MPA sensitivity (MPA$^S$) and is indicative of TSS upstream shifting. **(B)** Schematic illustration of *imd2Δ::HIS3* reporter and its conditional transcripts. The *IMD2* ORF is replaced by the *HIS3* ORF; when the downstream *IMD2* TSS is used constitutively, His3 product will be produced to support yeast growth on SC-His medium. Expression of *HIS3* will confer a His$^+$ phenotype and is indicative of TSS downstream shifting. **(C)** Schematic illustration of *gal10Δ56* reporter and its conditional phenotypes[77]. Deletion of *GAL10* polyadenylation signal at p(A) site interferes with *GAL10* 3'-end formation and *GAL7* initiation, which results in toxicity phenotypes when galactose is present. Transcriptional mutants that could increase *GAL10* 3'-end formation or *GAL7* initiation will suppress the toxicity and display GAL$^R$ phenotypes. **(D)** Schematic illustration of *lys2-128∂* reporter and its conditional phenotpyes[77]. Insertion of a Ty transposon into *LYS2* block Pol II readthrough and creates a short non-functional transcript, resulting in lysine auxotrophy. Certain

mutants that allow expression of *LYS2* from a cryptic site will support yeast growth on SC-Lys plate and confer a Spt$^-$ phenotype.

### 2.3.2 Quantification of TSS usage at *ADH1*

*ADH1* is broadly used as a model gene for TSS studies in *S. cerevisiae.* It contains two major "A" TSSs that are 82 and 92 nucleotides downstream of TATA element (from the last A of TATA box), shown as TATAAATA($N_{82}$)**A**agctatacc**A (Figure 13A)**. Using primer extension, transcription products of these two TSSs appear as bands of differing mobility on denaturing polyacrylamide gels **(Figure 13B, lane 1)**. Other positions show minor TSS usage. In most published papers, the two major starts primarily are compared qualitatively. To quantitatively compare all TSS usage at *ADH1*, we divide the initiation region of the *ADH1* promoter into six bins from upstream to downstream, in which bin three and five each contain one of the major *ADH1* mRNA isoforms **(Figure 13A, B)**. In order to compare a mutant's TSS usage to the WT TSS distribution, each WT bin signal is normalized to total signal from all bins and subtracted from the relevant normalized mutant bin signal **(Figure 13C, D)**. A negative value indicates that the mutant has relatively lower usage for that particular bin of starts and is represented as a bar below x-axis **(Figure 13D)**. A positive value indicates a relatively higher usage for that particular bin of starts and is represented as a bar above x-axis **(Figure 13D)**. For example, Pol II GOF allele E1103G increases relative TSS usage at upstream minor sites and decreases relative TSS usage at the downstream major site. Because of the dramatic effect of E1103G on TSS usage, the change of TSS usage can be visually detected on primer extension gel **(Figure 13B, lane 2)**. However, only quantitative comparison at fine bins could tell us how much TSS shift is made between WT and the mutant **(Figure 13D)**.

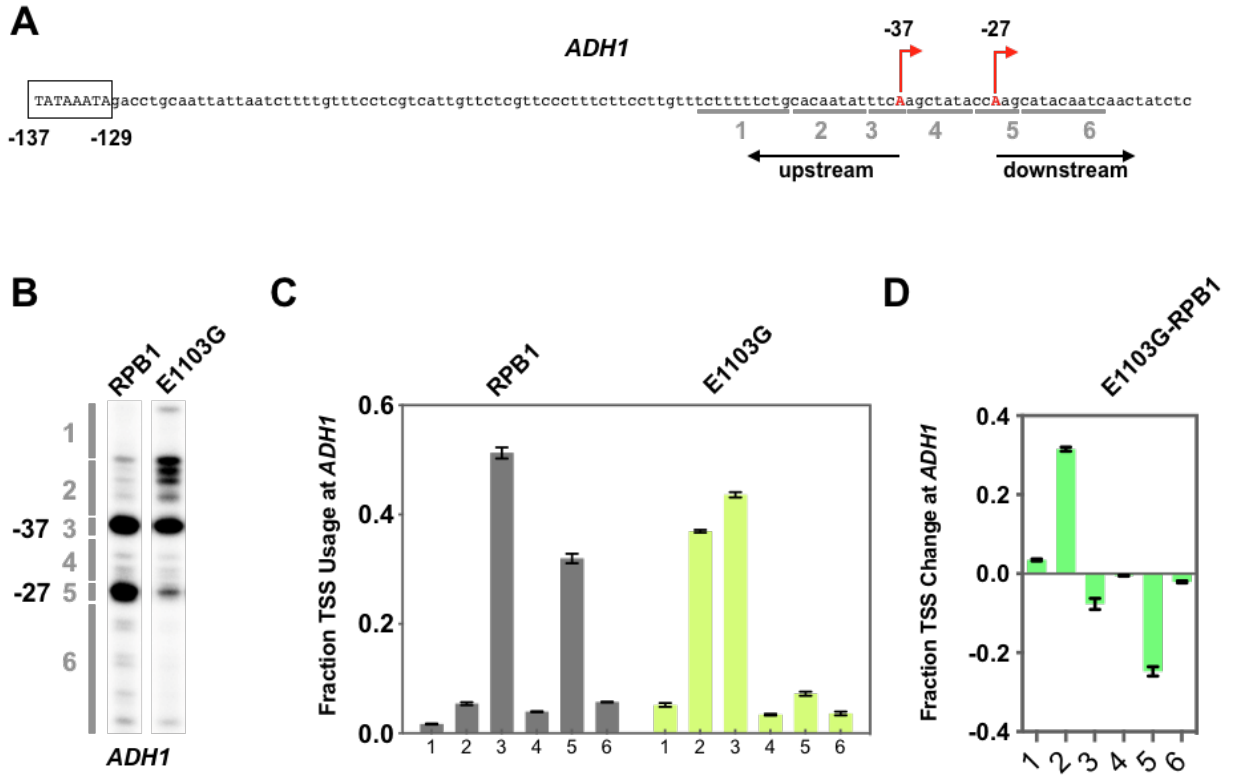



**Figure 13 Quantification of TSS usage at *ADH1***

**(A)** Schematic showing *ADH1* promoter architecture. The *ADH1* promoter contains two commonly used TSSs (red letters), which are 37nt (-37) and 27nt (-27) away from start codon respectively. The *ADH1* TATA-box is 129nt upstream of the start codon. Nucleotides at or upstream of -37 TSS are designated as upstream sites in our research, whereas nucleotides at or downstream of -27 TSS are called downstream sites. The *ADH1* initiation region is divided into six bins as indicated for quantification of primer extension data. **(B)** Example of primer extension products for detecting TSS usage at *ADH1*. WT yeast primarily use two major TSS in bin 3 and 5, while minor TSSs are used at low level. The Pol II catalytic activity GOF allele E1103G shifts *ADH1* TSS distribution upstream through increasing TSS usage at normally poorly used upstream minor sites, mainly in bin2. **(C)** Distribution of WT or mutant TSS signal in six *ADH1* bins. Each bin is quantified as the percentage of the total signal in the lane and displayed in a bar graph. The Pol II E1103G mutant increased TSS usage percentage in bin 2 from ~5% to ~37%, and relatively decreased TSS usage at bin 5 from ~32% to ~7%. **(D)** Quantitative showing the changes of TSS usage in E1103G compared to WT. Change of TSS usage at each bin is measured by subtracting the calculated percentage

of WT signal from that of the mutant strain. By subtracting the correspondent values in each bin of in bar graph C. A negative value is showed as a downward bar to indicate the decreased usage compared to the WT, whereas positive value shown in upward bar indicates an increased usage relative to the WT. The more change of TSS usage is made, the more deviates the bar from the x-axis will be observed. Pol II E1103G mutant increases TSS usage at bin 2 about 37%-5%=32%, and decreased in bin 5 about 7%-32%=25%.

### 2.3.3 Existing *SSL2* alleles show transcriptional growth defects and distinct TSS usage at *ADH1* promoter

Among five existing *ssl2* mutants I tested in this study: *ssl2-rtt* (*ssl2* E556K), *ssl2-DEAD* (*ssl2* V490A/H491D), *SSL2-1* (*ssl2* W427L), *ssl2-508* (*ssl2* H508R), *rad25-ts24* (*ssl2* V552I/E556K), **(1)** *ssl2 DEAD* and *ssl2-508* both contain a mutation at the helicase domain and show different levels of MPA sensitivity, with *ssl2 DEAD* exhibiting a strong $MPA^S$ phenotype and *ssl2-508* exhibiting a weak $MPA^S$ phenotype, respectively **(Figure 14)**. Both alleles hypothetically shift TSS distribution upstream as predicted by the *IMD2* promoter. **(2)** *SSL2-1* also contains a mutation in the helicase domain and shows a $His^+$ phenotype, which hypothetically results from downstream shifts in TSSs. **(3)** *ssl2-rtt* and *rad25-ts24* show temperature-sensitive phenotypes, however, no transcription-related phenotypes are observed. Because of the absence of $His^+$ or $MPA^S$ phenotype, we predicted that these two alleles do not shift TSS usage. Notably, there is no $Spt^-$ phenotype observed among these five existing *ssl2* alleles. This observation is in contrast to our lab's previous study of Pol II alleles, where the $Spt^-$ phenotype was observed in a subset of $MPA^S$ alleles.

We next examined TSS usage of the above five existing *ssl2* mutants at the model gene *ADH1*. As predicted from the phenotypes observed, **(1)** mutants of *ssl2-DEAD* and *ssl2-508* showed upstream shifts to *ADH1* TSS distribution **(Figure 14B, lane 3 and 5)**. However, we

observed that these two upstream shifting *ssl2* alleles behaved differently from Pol II and other GTFs allele that shift TSS distribution upstream, but through increased usage of upstream minor TSSs **(compare Figure 13B lane 2 and Figure 14B lane 3/5)**. **(2)** Consistent with the prediction of genetic reporter *imd2Δ::HIS3*, the $His^+$ *SSL2-1* mutant shifts the overall TSS distribution downstream in our precision quantification TSS analysis **(Figure 14A, B, C)**. However, in a previous study of *ssl2* alleles, it was concluded that *SSL2-1* had no obvious effects on TSSs distribution when comparing usage of just the two major starts[56]. In our analysis, we found that *SSL2-1* increases the downstream start site utilization at bin 5 and 6 and relatively decrease TSS usage at the upstream major site **(Figure 14C, panel 3)**. Additionally, we noticed that *SSL2-1* allele shows very similar TSS usage to Pol II LOF mutants at *ADH1*[75]. **(3)** As predicted, two other mutants *ssl2-rtt* and *rad25-ts24*, had no obvious effects on TSSs utilization at *ADH1* **(Figure 14B, lane 2 and 6; 14C)**.

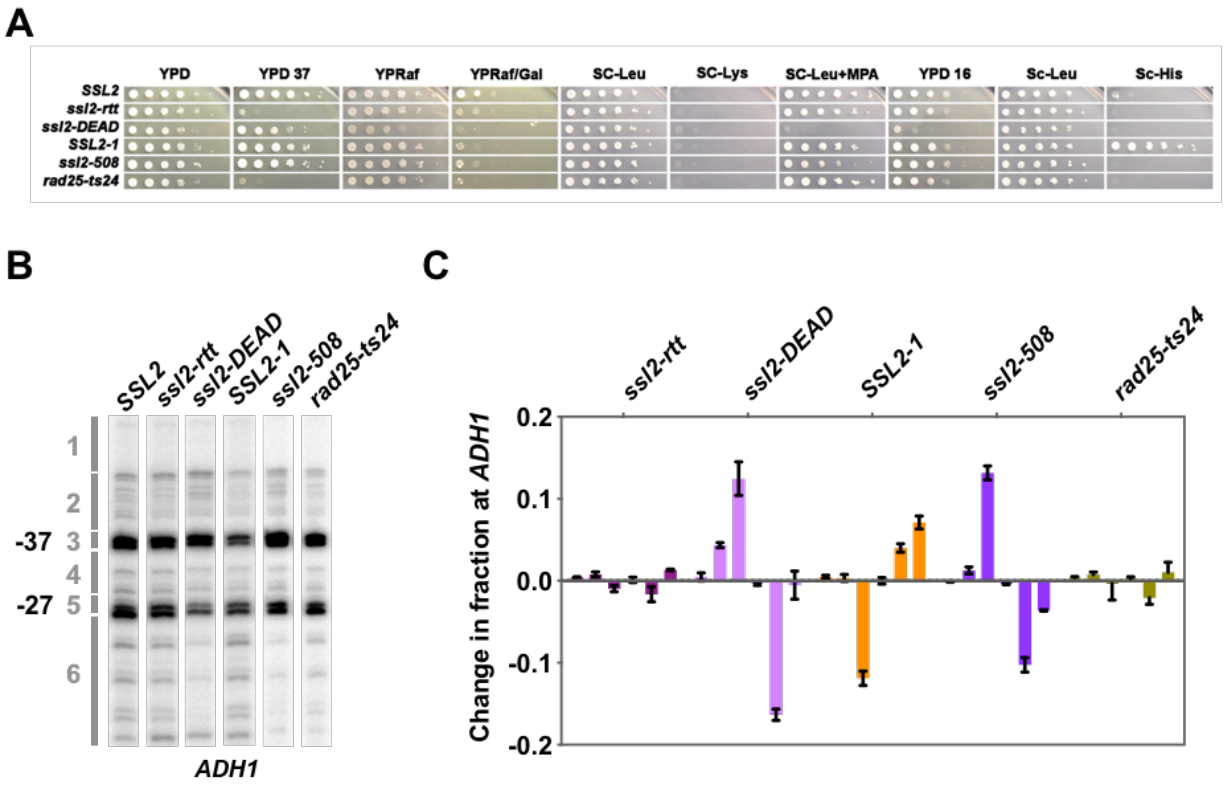



**Figure 14 Existing *ssl2* mutants show transcription-related phenotypes and distinct TSS usage at *ADH1***

**(A)** Transcription-related growth phenotypes of the five existing *ssl2* mutants tested by spot assay. Genetic reporters are described as in the main chapter and test medium are used as described in Materials and Methods. *ssl2-DEAD* allele shows a strong $MPA^S$ phenotype. *ssl2-508* exhibites mild $MPA^S$ phenotype. *SSL2-1* shows a $His^+$ phenotype. *ssl2-rtt* and *rad25-ts24* show no transcription-related phenotypes. **(B)** Primer extension detection of TSS usage at *ADH1*. *ssl2-DEAD* and *ssl2-508* with $MPA^S$ phenotypes shift TSS distribution upstream by decreasing usage at downstream sites. *SSL2-1* with $His^+$ phenotype shifts TSS distribution downstream by increasing downstream site usage. Mutant *ssl2-rtt* and *rad25-ts24* without putative transcriptional phenotypes show similar TSS usage as WT. **(C)** Quantitative change of TSS usage at *ADH1* promoter. *ssl2-DEAD* and *ssl2-508* dramatically decrease TSS usage at downstream bin 5 without activating upstream sites. *SSL2-1* increase TSS usage at downstream bin 5 and 6. Consistently, *ssl2-rtt* and *rad25-ts24* show little change of TSS usage in each bin.

### 2.3.4 Human disease-related and RED motif mutants

We additionally constructed and tested human disease-related *XPB* mutations in the yeast system, together with RED motif mutations. The amino acids altered by these mutations are located in the region that is conserved between XPB and Ssl2, which makes the dissection of their functions in yeast meaningful. Among four human disease-related mutants we tested, F156S, T176P, Q592* and Y750*, one confers lethality, Q592*, which creates a C-terminally truncated Ssl2 protein **(Figure 15A)**. T176P causes little if any growth defects and no $MPA^S$ or $His^+$ phenotypes; therefore, we hypothesize that this allele is not likely to alter TSS usage **(Figure 15B)**. In contrast, F156S confers a mild $MPA^S$ phenotype and shifts TSS distribution upstream at *ADH1* **(Figure 15B-D)**. Mutant Y750*, which mimics a disease related C-terminally truncated protein, shows a mild to moderate level $MPA^S$ phenotype and shifts TSS distribution upstream at *ADH1* **(Figure 15B-D)**. In addition, among 4 RED motif mutants we tested, R519A, E520A, E521A, and the triple mutant that contains all three mutations of R519A E520A E521A, all show lethal phenotypes. This is different from RED mutants' behavior observed in *Archaeoglobus fulgidus*. Instead of a lethal phenotype, RED mutants in *A. fulgidus* are alive and have dramatically reduced helicase activities[64]. The lethal phenotypes of RED motif substitutions in *ssl2* mutants revealed their essential roles in maintaining yeast cell functions.

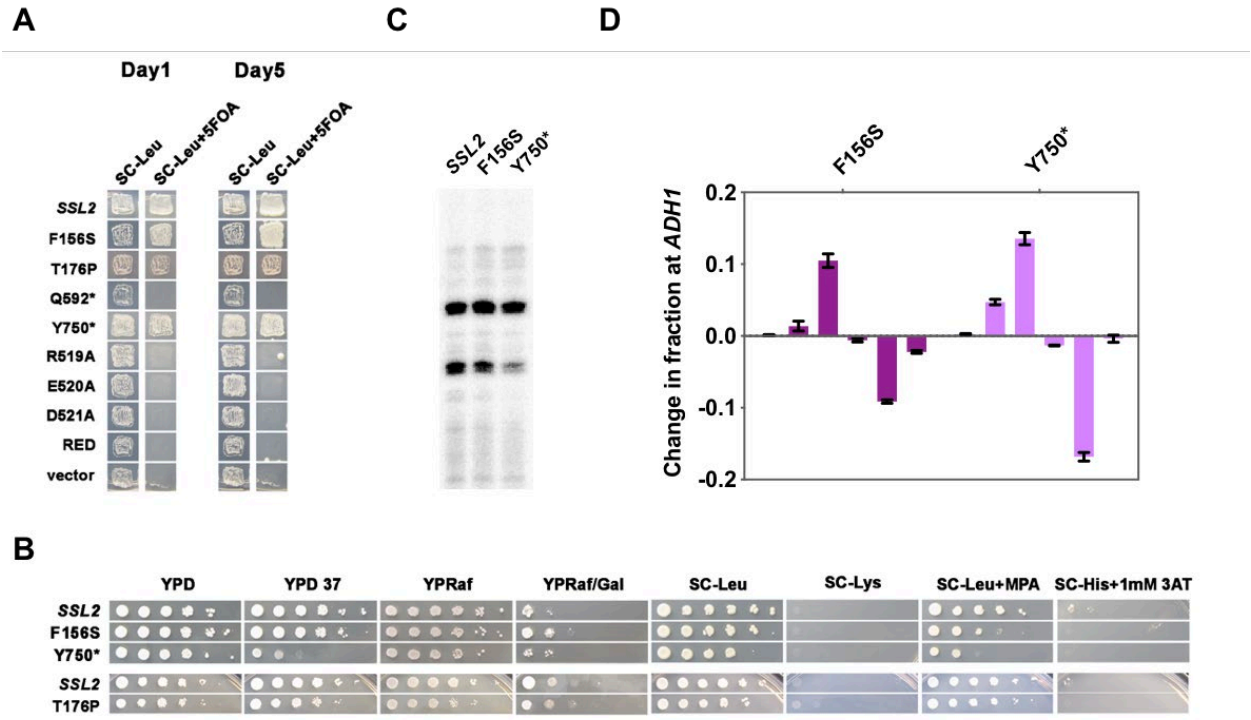



**Figure 15 Some human-disease related *ssl2* homologous mutants show transcription-related phenotypes**

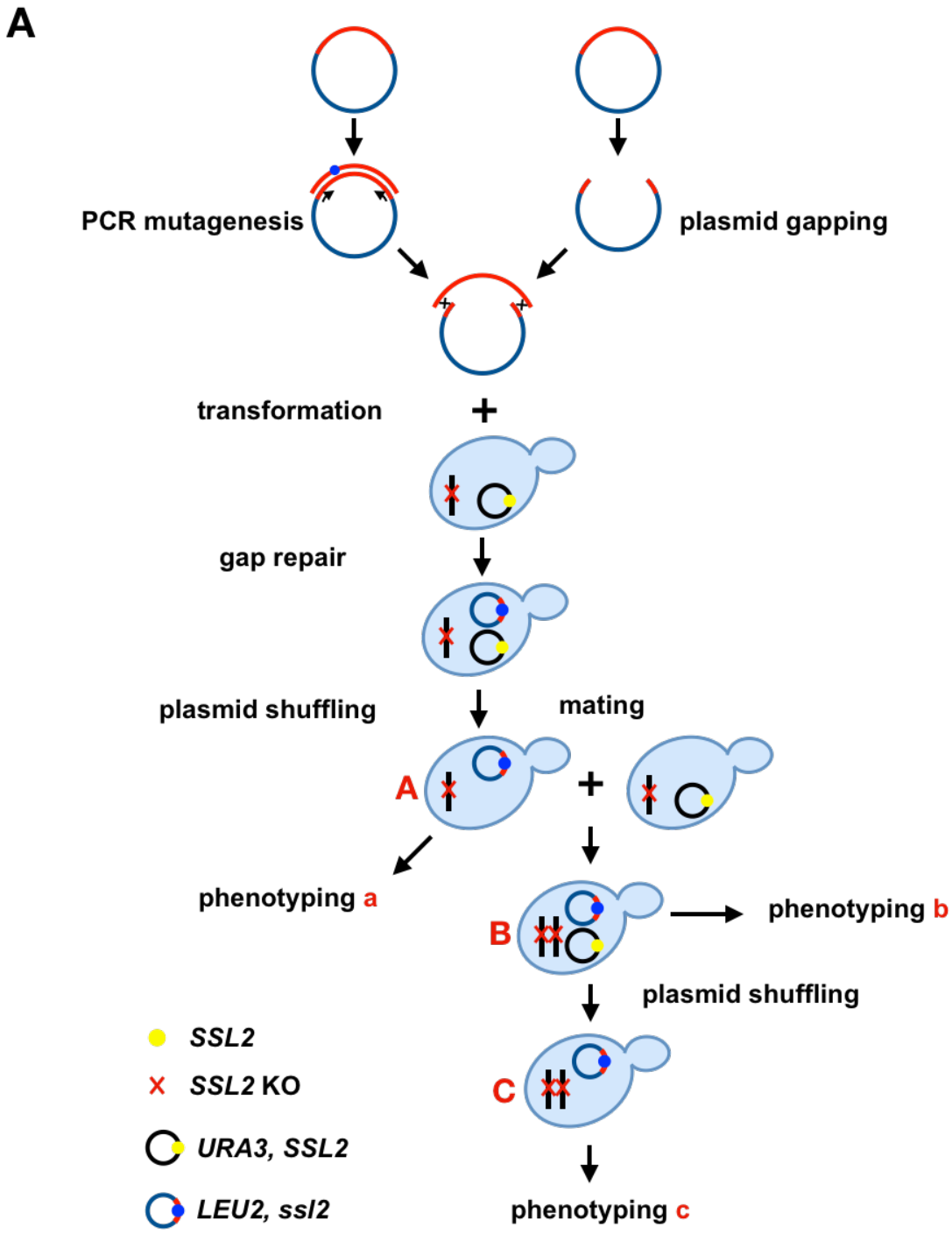
**(A)** Plasmid shuffling and patch assay to test five existing *ssl2* mutants' abilities in supporting yeast growth. Allele *ssl2* Q592* and all the RED motif alleles, R519A, E520A, E521A and R519A/E520A/E521A show lethality phenotyes. **(B)** Transcriptional related growth phenotypes of the *ssl2* mutants F156S, Y750* and T176P tested by spot assay. Both *ssl2* F156S and *ssl2* Y750* show $MPA^S$ phenotypes. Mutant *ssl2* T176 is slightly defective in growth, however, it doesn't show $MPA^S$ or $His^+$ phenotypes. **(C)** Primer extension to detect TSS usage of *ssl2* F156S and *ssl2* Y750* at *ADH1*. Both *ssl2* F156S and *ssl2* Y750* shift TSS distribution upstream through decreasing usage at downstream sites. **(D)** Quantitative change of TSS usage at *ADH1* promoter. Alleles of F156S and Y750* both decrease TSS usage at downstream bin 5 and relatively increase TSS usage at bin 3. Y750* additionally increases relative TSS usage at bin 2.

### 2.3.5 Gap-repair strategy for *ssl2* mutant screening

*ssl2* mutants were created by PCR-based random mutagenesis coupled with a gap repair method as previously described **(Figure 16)**[77]. Briefly, mutation of *SSL2* (*ssl2**) was accomplished

by standard PCR reactions using Taq polymerase (New England Biolabs). *ssl2** PCR products were then transformed into yeast along with a restriction digested, linearized pRS315 *SSL2 LEU2* plasmid with most of the wild-type *SSL2* sequence removed by restriction digest. $Leu^+$ transformants were selected. Homologous sequences on each end of the *ssl2** PCR products and the gapped *SSL2* vector allows homologous recombination, resulting in a library of gap-repaired plasmids containing potential *ssl2** alleles. Since *SSL2* is essential, these yeast cells are pre-transformed with a pRSII316 *SSL2 URA3* plasmid to support growth, while the genomic *SSL2* is deleted to allow plasmid *SSL2* alleles to exhibit phenotypes. After gap repair, cells retaining pRSII316 *SSL2 URA3* plasmids were killed by replica-plating transformants to medium containing 5FOA. Yeast cells were then plated on YPD media for growth and replica plated to a variety of media to screen for mutant that have transcription-related or conditional phenotypes. Plasmids from yeast mutants were recovered and transformed into bacteria cells for amplification, followed by sequencing to identify mutations. These mutant candidates were additionally mated with yeast cells that contain a WT *SSL2 URA3* plasmid to create diploid strains and perform phenotyping again to determine the dominant or recessive type of *ssl2* mutations. Plasmid shuffling on diploid strains was performed by adding 5FOA on the medium so that presumably only one copy of mutagenized *SSL2* was kept. This was followed by an additional phenotyping to determine if the mutant phenotype is plasmid linked or not. In total, there will be three phenotyping steps (a, b, c) for mutants at three different stages (A, B, C) harboring three different genotypes. The first stage mutant A is a haploid containing a putative *ssl2* LEU2* plasmid and *ssl2*Δ in the genome. The second stage mutant B is a diploid with a putative *ssl2* LEU2* plasmid, an *SSL2 URA3* plasmid and *ssl2*Δ*/ssl2*Δ in the genome. The third stage mutant C is a diploid with one plasmid, the putative *ssl2* LEU2* plasmid, and *ssl2*Δ*/ssl2*Δ in the genome. Comparison of the phenotypes between A

and B determines if the mutation is dominant or recessive, and the comparisons of the phenotypes between A, B and C determine, for recessive mutants, if the mutant phenotype is plasmid linked, see illustration in **Figure 16B**.



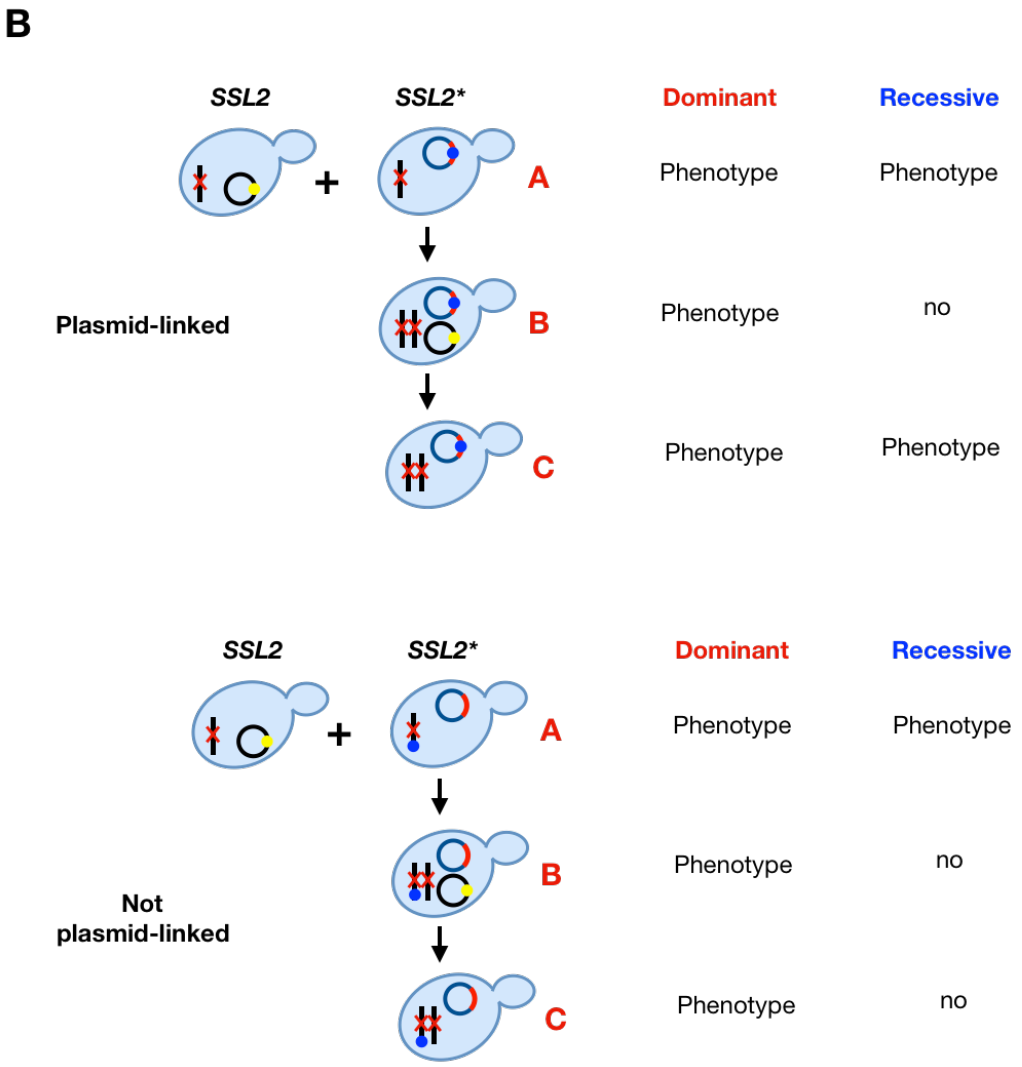



**Figure 16 Gap-repair strategy for *ssl2* mutant screening**

**(A)** Gap-repair strategy for screening transcription-related *ssl2* mutants. A *LEU* plasmid containing *SSL2* sequence was used as a template to create mutagenized *ssl2* PCR fragments. Another copy of *SSL2* on a *LEU2* plasmid was digested by restriction enzymes to create flanking sequences for recombination. The mixture of PCR and digested plasmid were then transformed into *MATa* yeast containing a WT *SSL2 URA3* plasmid. The mixed PCR fragment and digested plasmid will be recombined *in vivo* and results in $Leu^+$ yeast transformants. These transformants were then cultured on SC-Leu+5FOA plate to select for cells have lost the WT *SSL2 URA3* plasmid. Yeast colonies grow at this step represent potential haploid mutants for genetic screening and identification of candidate mutants. After first round of phenotyping (phenotyping a), colonies showing transcription related phenotypes (mutant A) were struck on SC-Leu plates for single colony purification and rescreening. These single colonies were then mated with

*MATα* yeast cells containing a WT *SSL2 URA3* plasmid to create diploid strains (mutant B). These diploid strains were also phenotyped (phenotyping b) to determine dominant/ressiveness of apparent phenotypes. Then, plasmid shuffling was applied by adding 5FOA on the culturing plates to get diploid yeast retaining only the *LEU2*-marked *ssl2** alleles (mutant C). This final diploid mutant was tested by phenotyping again (phenotyping c) to determined if the mutant phenotype is plasmid-linked. **(B)** Determination of dominant/recessive mutation and the plasmid-linked phenotypes. Mutant A is a haploid containing a putative *ssl2* LEU2* plasmid with a chromosomal deletion of *SSL2* (*ssl2*Δ). Mutant B is a diploid with a putative *ssl2* LEU2* plasmid, a *SSL2 URA3* plasmid and two copies of WT *SSL2* deleted genomes. Mutant C is a diploid with a putative *ssl2* LEU2* plasmid and two copies of WT *SSL2* deleted genomes. Comparison of the phenotypes between A and B determines if the mutation is dominant, showing phenotypes in both mutant A and B, or recessive, exhibiting phenotypes in mutant A but not in B. Comparisons of the phenotypes between A, B and C determine if the mutant phenotype is plasmid linked or not. Dominant mutant will show phenotypes in A, B and C no matter plasmid-linked or not. Recessive mutant will show phenotypes in A and C, but not in B if its plasmid-linked. In contrast, recessive mutant will exhibit phenotypes in A, but not in B and C if it is not plasmid-linked.

### 2.3.6 Genetic screening identified *ssl2* mutant with transcriptional defects

We screened for *ssl2* mutants with transcription-related phenotypes by "gapping strategy" as illustrated in 2.3.5. The transcription-related phenotypes include Spt$^-$, His$^+$, MPA$^S$ and Gal$^R$, and the temperature sensitive phenotype Tsm$^-$, as described in 2.3.1. In total, we screened 28685 yeast colonies, counted by BioRad Gel Doc. In the first-round screening, 853 mutant candidates were found. Among these, mutant candidates showing relatively strong transcription-related or temperature sensitive phenotypes were selected and reanalyzed by phenotyping, after which 273 candidates were selected for dominant/recessive assessment. We observed that dominant mutations are found for Spt$^-$, His$^+$ and Gal$^R$ phenotypes in tested mutants; however, there were no dominant mutants found for Tsm- and MPA$^S$ phenotypes, consistent with either recessive loss of

function mutations or the nature of the phenotype (sensitivity) or both. The recessive mutants with plasmid-linked phenotypes (130 in total) were all selected for plasmid recovery and sequencing. We additionally picked 19 mutants that had dominant Spt$^-$ phenotypes for plasmid recovery and sequencing. Our sequencing result for 149 selected mutant candidates showed that each of them contain 1-5 mutations. Some mutations were identified multiple times. 42 each contained a unique single amino acid substitution and were selected for further analysis. These 42 *ssl2* single mutants are listed in **Table 8.**

**Table 8 *ssl2* mutants with single amino acid substitutions**

| Number | Amino acid substitution | Number | Amino acid substitution | Number | Amino acid substitution |
|---|---|---|---|---|---|
| 1 | I170T | 15 | K328N | 29 | D522V |
| 2 | L180S | 16 | E340G | 30 | K523N |
| 3 | S185Y | 17 | N345T | 31 | N528I |
| 4 | S201Y | 18 | D346G | 32 | F529L |
| 5 | I218T | 19 | N349D | 33 | E537G |
| 6 | L225P | 20 | K357E | 34 | W558R |
| 7 | V226D | 21 | K357R | 35 | D610A |
| 8 | N230D | 22 | K372E | 36 | R636C |
| 9 | N230I | 23 | G382V | 37 | N642D |
| 10 | N230S | 24 | I383T | 38 | L664S |
| 11 | Y232D | 25 | T402I | 39 | Q672R |
| 12 | L246P | 26 | K443E | 40 | E715G |
| 13 | V268A | 27 | V473D | 41 | S719P |
| 14 | N320Y | 28 | F498L | 42 | L832* |

### 2.3.7 *ssl2* mutants fall into two major classes in terms of *in vivo* conditional phenotypes

For 42 *ssl2* single mutants we identified from the screening, we performed spot assay to confirm and test the strength of their transcriptional phenotypes **(Figure 17)**. These single mutants can be broadly classified into two major classes in terms of their *in vivo* phenotypes: one class is defective for the induction of the *IMD2* gene and results in an $MPA^S$ phenotype, the other class shows constitutive expression of the *imd2::HIS3* fusion reporter gene and confers a $His^+$ phenotype. Other transcription-related phenotypes, $Spt^-$, $Gal^R$ and $Tsm^-$, were observed as subsets of these two major classes. The primer extension results show that these two major classes of *ssl2* mutants show expected TSS shift as predicted by the genetic reporters, with all the tested $MPA^S$ alleles shifting TSS distribution upstream and all $His^+$ alleles shifting TSS distribution downstream **(Figure 18)**.

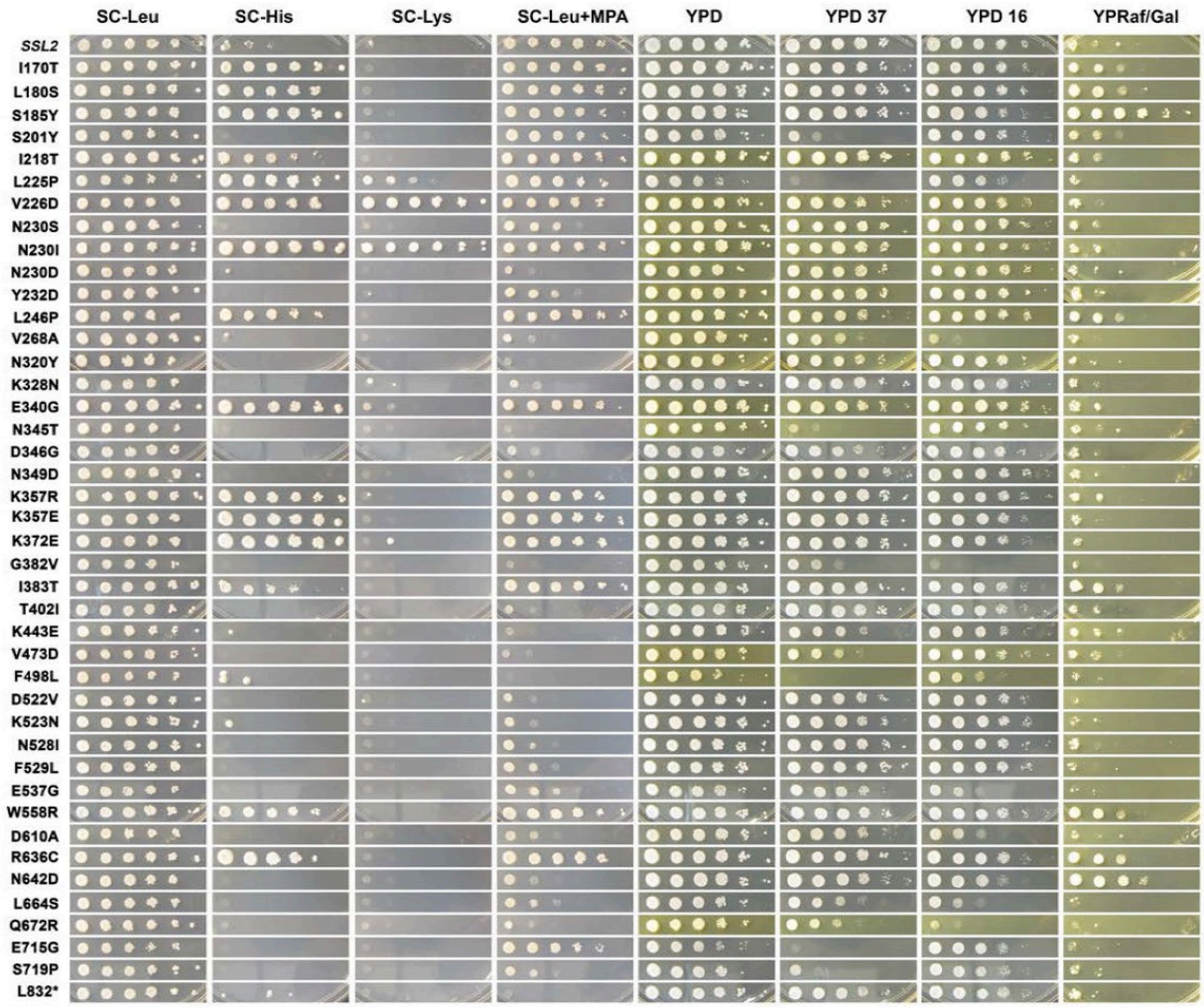



**Figure 17 Spot assay for transcription-related phenotypes of 42 *ssl2* single mutants**

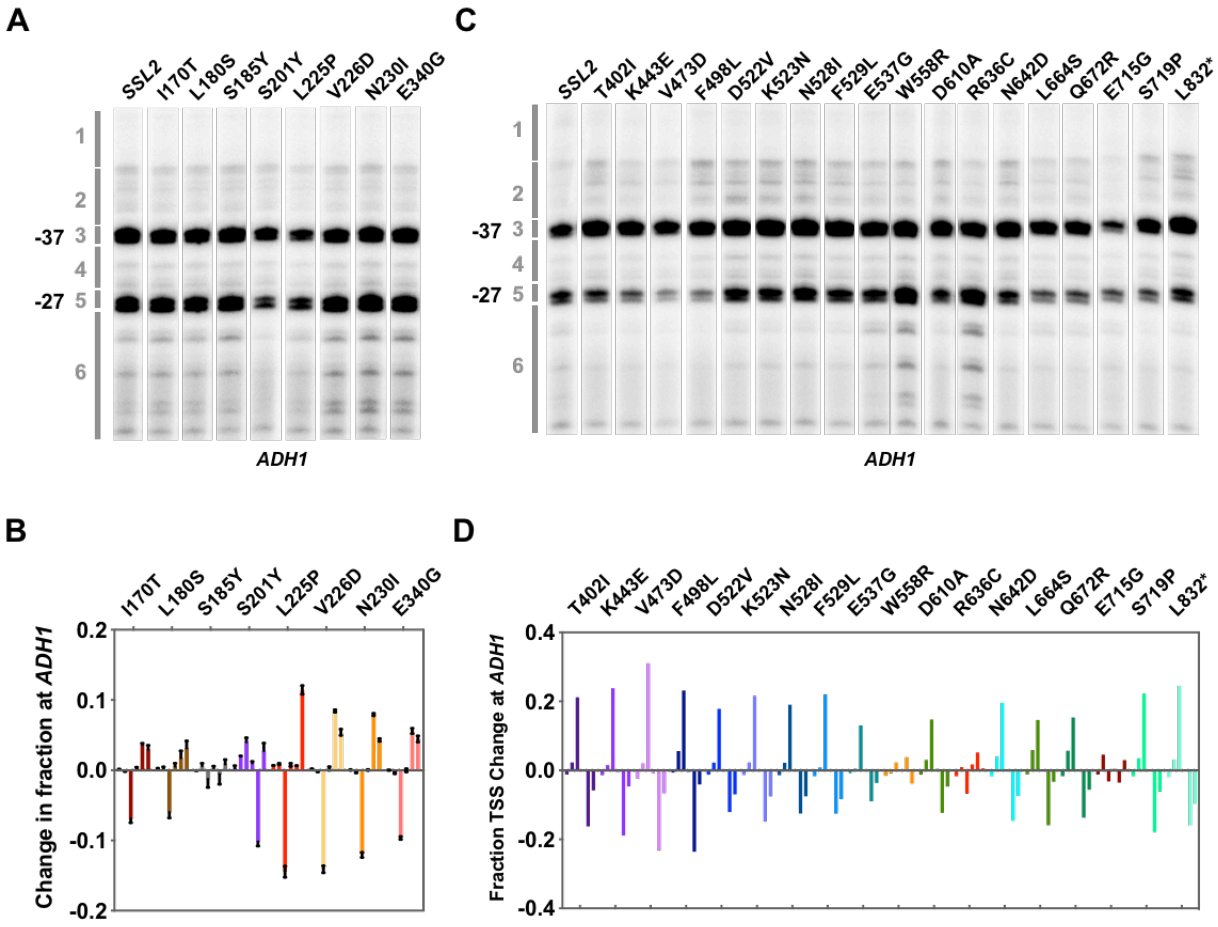



**Figure 18 TSS usage of some representative *ssl2* single mutants**

**(A)** Primer extension detected TSS usage for selected *ssl2* alleles at *ADH1*. **(B)** Quantification of TSS usage for selected *ssl2* alleles at *ADH1*. **(C)** Primer extension detected TSS usage for some other *ssl2* alleles at *ADH1*. Since our primer extension data indicated that all the tested *ssl2* mutants showed genetic reporter predicted TSS shifts in (A, B), to survey as many mutants as we could but avoid a large number of primer extension reactions, we performed primer extension on additional *ssl2* mutants but used only one biological replicate. As expected, all additionally tested *ssl2* $MPA^S$ mutants shift TSS upstream and $His^+$ mutants shift TSS downstream. **(D)** Quantification of TSS usage of *ssl2* mutants in (C).

### 2.3.8 *ssl2* mutant screening found new functional alleles with distinct transcriptional phenotypes

We next mapped mutation sites of our screened single mutants onto the Ssl2 primary structure. We observed that the mutations causing $MPA^S$ phenotypes alter amino acids distributed across the protein, with a high frequency at the two-helicase domains **(Figure 19A)**. In contrast, mutations related to $His^+$ phenotypes alter amino acids clustered at the N-terminal side of the protein. In addition, the $Spt^-$ phenotype is a new phenotype that has not been observed in the existing *ssl2* mutants, and is found exclusively linked to $His^+$ mutants. This link between $His^+$ and $Spt^-$ in *ssl2* mutants is different from what we have observed for Pol II mutants. In our previous studies of Pol II mutants, $Spt^-$ phenotype was observed for known Pol II catalytic center TL GOF alleles, and many of these $Spt^-$ mutants are also strongly $MPA^S$. Furthermore, TL LOF alleles confer $His^+$ phenotype on this genetic reporter. However, in the screening of *ssl2* alleles, none of the $Spt^-$ mutants were also $MPA^S$. We hypothesize this new $Spt^-$ class represents *ssl2* alleles with distinct functions. Indeed, these $Spt^-$ mutations were located within a newly identified LOCK-N region (residues 88-285 for Ssl2), which was shown to crosslink with other TFIIH subunit, Tfb3, Tfb6 and Ssl1 **(Table 4)**.

Mapping of *ssl2* mutations to the three dimentional structure of Ssl2 revealed that some TSS shifting mutations are located on the Ssl2 surface **(Figure 19B)**. Several interesting sites are found, including D610 located proximal to DNA. A substibution at this site, D610A, causes an upstream TSS shift **(Figure 19B, Front)**. In contrast, substitution of R636, which is also close to the DNA strand, to cysteine causes a downstream TSS shift. F498, which is located in the groove of the Ssl2 lobe 1 and facing the DNA strand, caused an usptream TSS shift when substituted with leucine and showed synthetic lethality with *sub1Δ*, *sua7-1* and *tfg2Δ146-180*, as discussed in

Chapter III. Additionally, a small patch of residues that has a high frequency of TSS shifting substitutions is found on the Ssl2 lobe 2 surface. These substitutions are from the helicase domain 1 and shift TSSs upstream, including D522V, K523N, N528I, F529L and E537G. In addition, substitutions of I383T and K372E are also found in this small patch region but shift TSS downstream **(Figure 19B, Right)**.

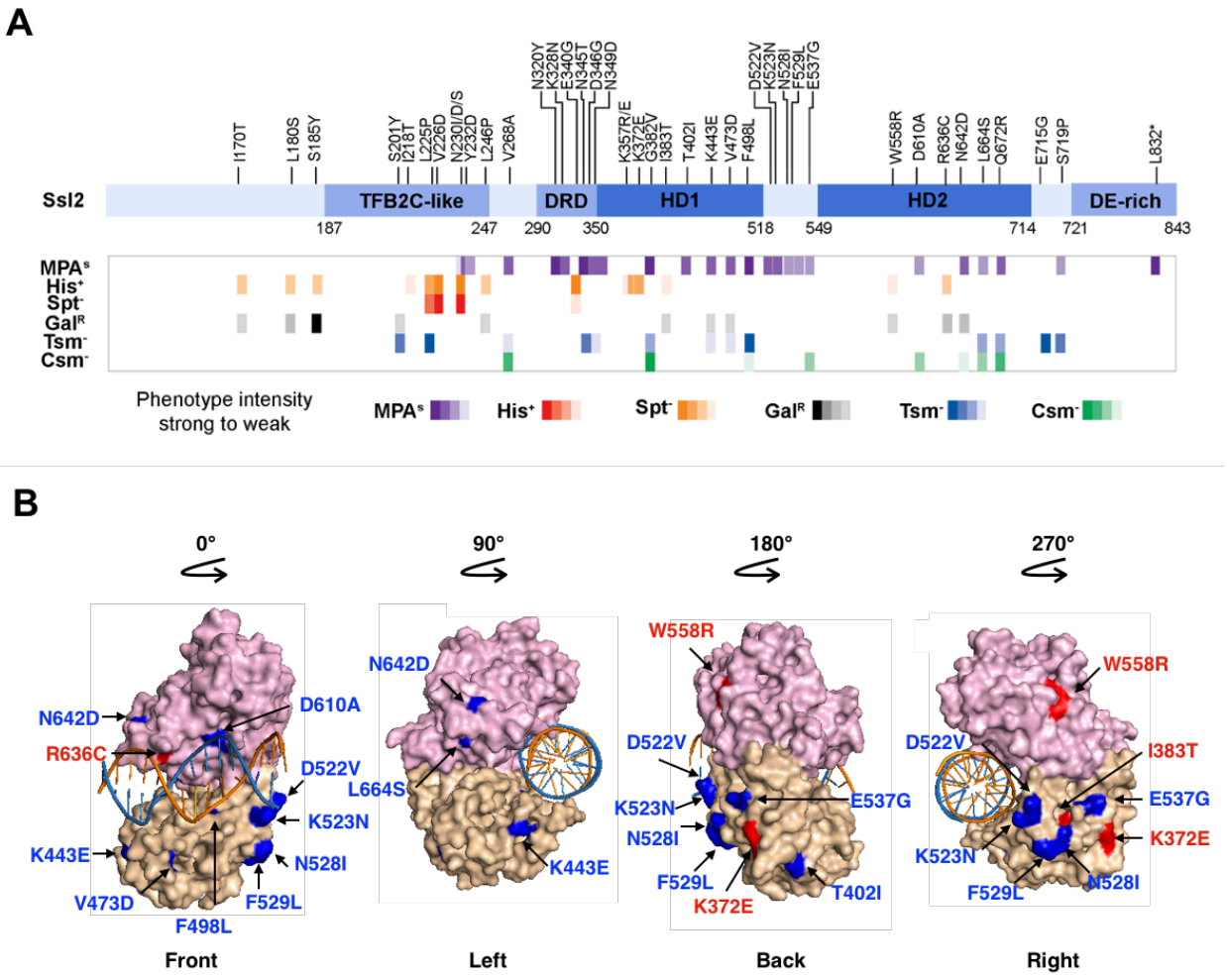



**Figure 19 Mapping newly identified TSS mutations to the Ssl2 protein structure**

**(A)** Mapping of identified *ssl2* single mutants and their transcriptional phenotypes to domains of the Ssl2 protein. Colored boxes represent strength of phenotypes with increased color intensity representing increse in phenotypic strength. *ssl2* mutants with MPA$^S$ phenotypes are distributed across the protein, especially within the helicase

domains. *ssl2* substitutions causing $His^+$ phenotypes were found mainly on the N-terminal side of the protein. $Spt^-$ alleles represent a subgroup of the $His^+$ alleles clustering tightly within the N-lock domain. $Gal^R$, $Tsm^-$ and $Csm^-$ alleles are subsets of $MPA^S$ or $His^+$ alleles without obvious clusering. **(B)** Mapping of *ssl2* substitutions to the Ssl2 protein structure. Protein structure used is *S. cerevisiae* PIC containg TFIIH (PDB 5OQJ). Two lobe-like modules of Ssl2 protein, which are N-terminal half (lobe 1) and C-terminal half (lobe 2), are separately colored as wheat-color and dusty pink. Amino acid substitutions causing $MPA^S$ phenotypes are labelled as blue and $His^+$ phenoytpes are showed as red.

## 2.4 DISCUSSION

Ssl2 is known to function in RNA Pol II transcription and NER. Mutations have been found in *SSL2* that affect transcription initiation and TSS usage. However, its function in TSS selection hasn't been extensively studied. We cloned existing *SSL2* mutations into our lab strains containing reporters for transcription-related phenotypes, tested them, and quantitatively measured their effects on TSS usage. These mutations include five previously found in yeast and suggested to be possibly defective in transcription initiation, *ssl2-rtt* (*ssl2* E556K), *ssl2-DEAD* (*ssl2* V490A/H491D), *SSL2-1* (*ssl2* W427L), *ssl2-508* (*ssl2* H508R), *rad25-ts24* (*ssl2* V552I/E556K). Four additional mutations we cloned into yeast are from human patients harboring disease-causative *XPB* mutations, F156S, T176P, Q592* and Y750*. We additionally created RED motif mutant *ssl2* alleles to test their effect on cell growth, which are *ssl2* R519A, *ssl2* E520A, *ssl2* E521A, and *ssl2* R519A E520A E521A. We found that the pattern of TSS usage by three different classes of *ssl2* alleles (alleles shifting TSSs upstream, alleles shifting TSSs downstream, and alleles having no obvious effect on TSS usage) correlated closely with *in vivo* growth defects. This

correlation suggests that roles of Ssl2 should be genetically dissectible and additional alleles might enhance this dissection. Thus, we performed a genetic screening for *ssl2* mutants.

Our genetic screening successfully identified *ssl2* mutants with transcription-related growth phenotypes and TSS defects. These *ssl2* alleles fall into two major classes in terms of transcriptional phenotypes and TSS usage tested at *ADH1* prompter. Representative *ssl2* alleles of the two major classes are *ssl2* N230D, which is MPA sensitive that fails to shift TSS downstream at *ADH1* promoter, and *ssl2* N230I, which shows a $His^+$ phenotype and shifts more downstream TSS than WT yeast **(Figure 20A, B)**. *rpb1* mutants also fall into two major classes regarding transcription phenotypes and shifting in TSS distribution, the two representatives are *rpb1* E1103G, which shows a $MPA^S$ phenotype and shifts TSS distribution upstream, and *rpb1* H1085Y that exhibits a $His^+$ phenotype and shifts TSS distribution to the downstream **(Figure 20C, D).** MPA-sensitivity phenotype correlates with upstream TSS distribution shift in both *ssl2* and *rpb1* upstream shifting alleles of N230D and E1103G, respectively. Similar to the observation with the existing *ssl2* $MPA^S$ alleles of *ssl2-DEAD* and *ssl2-508*, *ssl2* N230D shifts TSS distribution upstream by limiting TSS usage at the downstream sites but does not activate lowly used upstream sites, mainly decreasing TSS at bin 5 and 6 of *ADH1* promoter **(Figure 20B)**. However, *rpb1* E1103G shifts TSS distribution upstream through activating normally poorly used TSS at upstream bin 2 of *ADH1* promoter **(Figure 20D)**[75]. In addition, *ssl2* mutants showing the $Spt^-$ phenotype, a phenotype conferred by suppression of *lys2-128∂* insertion, and is found as a subset of strong $His^+$ alleles **(Figure 20A)**[107]. This connection of $Spt^-$ phenotype with $His^+$ is in contrast to what we have observed for $Spt^-$ alleles of Pol II mutants, which are generally a subset of $MPA^S$ alleles **(Figure 20C)**[75]. These differences between Pol II and *ssl2* alleles are consistent with their functions in activating TSS and suppressing *lys2-128∂* by distinct mechanisms. To explain these differences

and provide a framework with which to guide further experimentation, we proposed a Shooting Gallery model for promoter scanning and described in the next chapter.

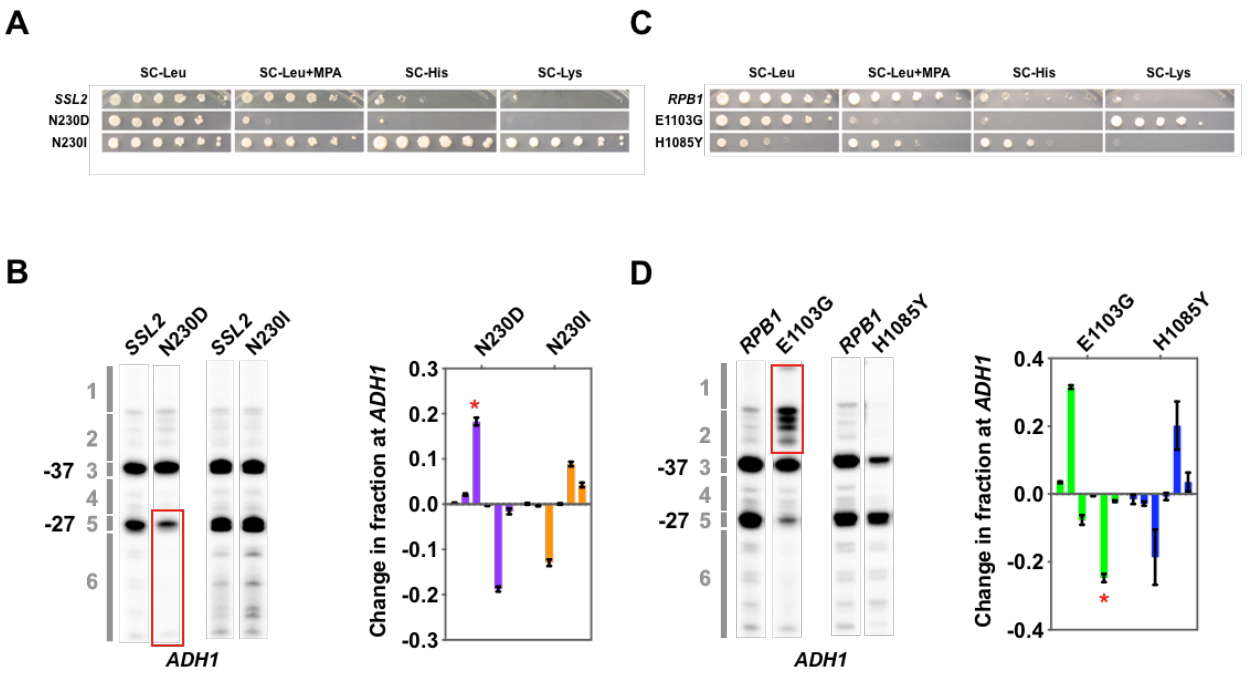



**Figure 20 Comparison of transcirptional phenotypes and TSS usage between *ssl2* and *rpb1* alleles**

**(A)** Spot assay showing *ssl2* mutants' transcription phenotypes. Allele N230D causes $MPA^S$ and is predicted to shift TSS distribution to the upstream region of the promoter. N230I is $His^+$, and $Spt^-$, and hypothetically shifts TSS distribution downstream. **(B)** Primer extension and quantification to show *ssl2* alleles' effect on TSS usage at *ADH1* promoter. N230D shifts TSS distribution to upstream side by decreasing downstream TSS usage at bin 5 and 6. N230I shifts TSS distribution to upstream side by increasing downstream TSS usage at bin 5 and 6. **(C)** Spot assay showing *rpb1* mutants' transcription phenotypes. Allele E1103G is $MPA^S$ and $Spt^-$, which has an increased Pol II catalytic activity and is known to shift TSS distribution upstream. H1085Y is $His^+$ and known to shift TSS distribution downstream. **(D)** Primer extension and quantification to show *rpb1* alleles' effect on TSS usage at *ADH1* promoter. E1103G shifts TSS distribution to upstream site by activating upstream TSS usage at bin 2. H1085Y shifts TSS distribution to dwonstream site by increasing TSS usage at bin 5 and 6.

# 3.0 GENETIC ANALYSIS OF SSL2 FUNCTIONS IN TSS SELECTION

## 3.1 INTRODUCTION

### 3.1.1 The Shooting Gallery model of TSS selection

During initiation, TFIIH opens double stranded DNA by translocation of downstream DNA into the Pol II cleft[5,108]. During initiation by scanning, as the single stranded template DNA passes the active center of polymerase, the TSS is located to start RNA synthesis[2]. Thus, the absolute utilization of a specific TSS is determined by Pol II catalytic activity, by the exposure time of a TSS to the Pol II active site, and by the ability of the DNA sequence to support initiation. These three aspects that determine the *efficiency of a TSS* can be modeled into a "Shooting Gallery" model, "where the rate at which a target passes, the rate of firing, and the strength of TSS in supporting transcription together contribute to the probability a target is hit" **(Figure 21A**, modified from the original statement of Craig Kaplan in)[101]. **(1)** Specifically, Pol II catalytic rate controls the incorporation of substrate nucleotides within a given time during transcription. This is an intrinsic property of RNA polymerase that is coupled with the efficiency of TSS usage, and is analogous to the firing rate in the Shooting Gallery model. **(2)** The exposure time is designated as "translocation rate" or "scanning rate", which is proposed to be facilitated by Ssl2's translocase activity and determines the rate of DNA passing the Pol II active site in the Shooting Gallery model. **(3)** Inr is the DNA sequence immediately surrounding and containing a TSS, and the sequence strength for initiation is analogous to the size of the shooting target in this model[109]. Some Inr sequences are intrinsically better for initiation than others, thus TSSs with various

strengths may be present within a scanning window. We can imagine that when DNA passes through Pol II active site, reduced Pol II activity *or* a faster scanning rate would both decrease the transcription efficiency at every TSS, by shifting competition between Pol II activity and template movement toward the latter. Conversely, increased Pol II activity or slower scanning rate would both increase the efficiency at individual TSS[101].

However, Pol II activity, scanning rate and Inr affect not only single site *efficiency* but also the overall TSS *distribution* when all TSS are considered within the scanning window. When we looked the overall TSS distribution in Pol II mutants with altered catalytic activity that is known to affect transcription efficiency, we observed polar changes to TSS distributions[101]. Distributions will also necessarily be shaped by an additional factor, the amount of Pol II available to initiate is expected to diminish downstream as cumulative initiation probability increases to one: **(4)** Pol II flux. Pol II flux is the relative number of polymerases encountering a given start site, which has a higher value at upstream TSS and a lower value at downstream TSS, resulting in reduced apparent usage at downstream position distinct from their inherent efficiencies **(Figure 21B, WT; Figure 21C, Pol II flux in red)**. As Pol II binds and scans from upstream to downstream, the apparent probability of initiation reaches one, and once it does so, no initiation will be observed beyond that point. The greater the amount of Pol II that enters initiation at upstream positions, the lesser the amount that will continue to scan to downstream positions. When Pol II has increased catalytic activity, *e.g.* Pol II catalytic activity GOF allele, upstream TSSs will be more efficiently used because Pol II gains the ability to use these sites, along with quickly reduced Pol II flux at downstream sites **(Figure 21D)**. In this allele, usage of downstream TSSs could be decreased, due to quickly reduced and insufficient Pol II flux at downstream sites, resulting in an efficiency curve with an increased slope and polar changes to TSS distribution **(Figure 21B, Pol II GOF; Figure**

**21D, green and black arrows as TSSs)**. Conversely, when Pol II has decreased catalytic activity, like Pol II catalytic activity LOF allele, TSSs at upstream sites will be less efficiently used, along with slowly reduced Pol II flux **(Figure 21E)**. Pol II activity from upstream insufficient usage will be reserved and applied to the downstream sites, resulting in apparently increased TSS usage at downstream sites and flattening and spreading of the TSS distribution (as demonstrated by an efficiency curve with decreased slope) **(Figure 21B, Pol II LOF; Figure 21E, blue and black arrows as TSSs)**. There are two additional factors that have not been previously discussed that will also have effects on overall TSS distribution; they are: **(5)** upstream and **(6)** downstream constraints for defining the scanning window. Studies suggest that very upstream TSSs close to the presumed location of PIC assembly show reduced transcription initiation. The physical basis for defining the upstream boundary of the scanning window has not yet been determined. We hypothesize, one constraint is the minimum space required for PIC assembly, however as yeast do not use TSSs positioned at distances used in other eukaryotes, there must be an additional constraint **(Figure 21C, gray curve)**. Moreover, we hypothesize that downstream constraints for defining the scanning window could either be TFIIH's processivity or Pol II flux **(Figure 21C, TFIIH)**. Previous single molecule studies suggested that TFIIH drives downstream scanning by extending an unpaired DNA region that is similar in length to the distribution of TSSs at yeast promoters[44]. Thus, the processivity of TFIIH scanning will control how far Pol II machinery can reach at downstream TSSs. Additionally, it's not clear if Pol II stops initiation due to exhausted flux prior to where TFIIH processivity terminates for any individual promoter. Nevertheless, we have these scenarios considered and come up with a "Cooperation" model to accommodate the above factors and their activities in promoter scanning. We propose that TSS distribution of a promoter is established by the cooperation of Pol II's catalytic activity and TFIIH's processivity

for activating TSS at promoter sites. We next use this model to predict Ssl2/TFIIH's function in TSS selection.

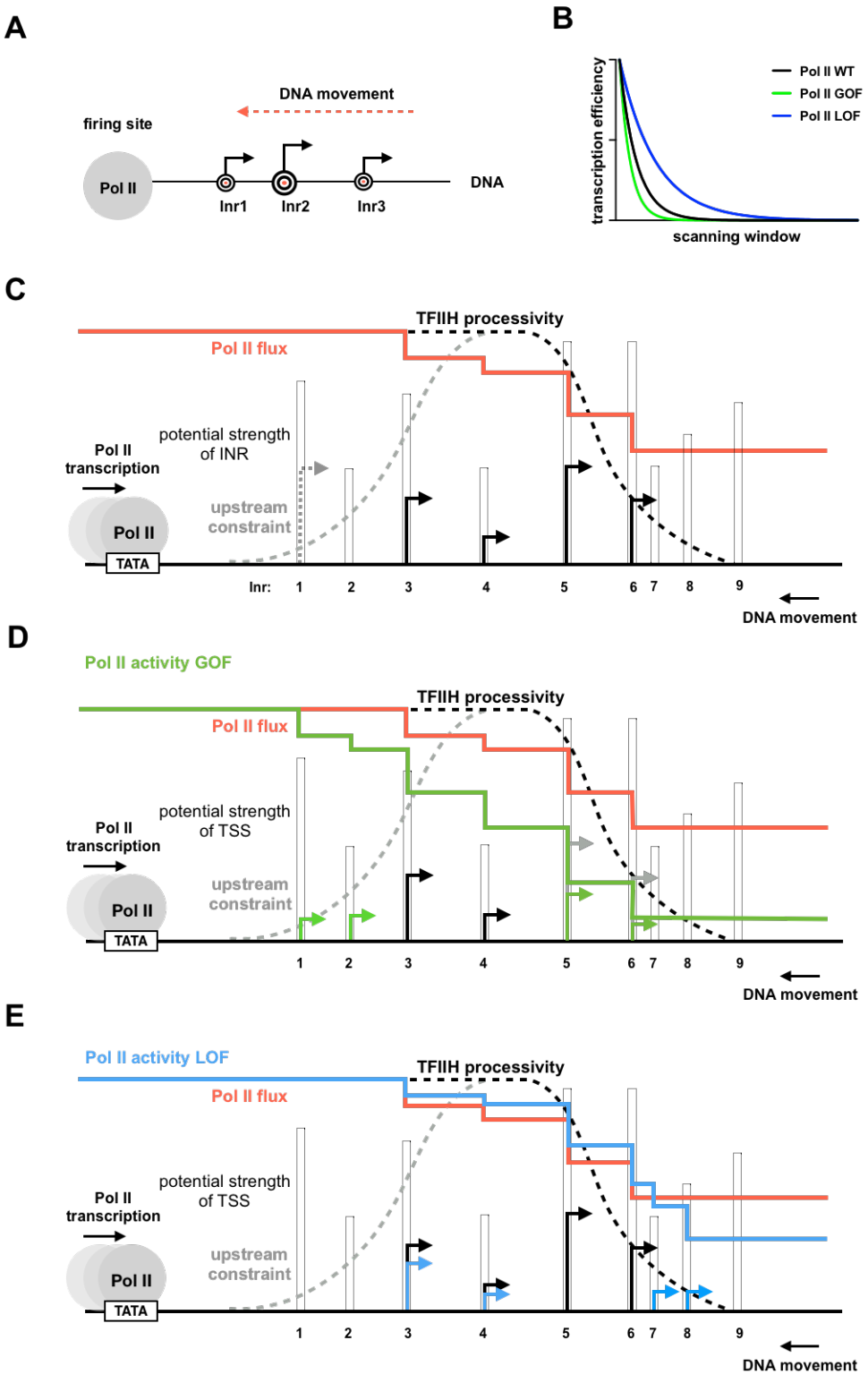


**Figure 21 The shooting gallery model**

**(A)** The "shooting gallery" model. The firing site mimics Pol II catalytic rate while scanning machinery will present Inr targets to the active site like ducks on a conveyor belt in a shooting gallery. The shooting target represents Inr that moves towards the firing site as DNA is scanned (moves relative to the Pol II active site). The probability a target (TSS) is hit is presumably determined by the rate of firing (scanning rate) and size of target (Inr strength). **(B)** Transcription efficiency *vs.* scanning window. Black curve: WT Pol II efficiency. Green curve: Pol II catalytic activity GOF mutant. Blue curve: Pol II catalytic activity LOF mutant. As Pol II binds and scans from upstream to downstream, successful initiation at upstream positions will cause lesser amount of Pol II continue to scan downstream positions and result in a decreased efficiency curve. When Pol II has increased catalytic activity, upstream TSSs will be more efficiently used and result in an efficiency curve with an increased slope. Conversely, when Pol II has decreased catalytic activity, reduced efficiency at upstream sites causes flattening of the efficiency curve. **(C)** The "Cooperation" model. Black line: schematic showing of a promoter DNA with TATA box at upstream. Arrows: TSS. Dotted box: potential strength of TSS. Red line: Pol II flux. Gray curve: upstream constraint. Black curve: WT TFIIH processivity. As Pol II binds and scans from upstream to downstream. The TSS distribution of a promoter window will be determined by Pol II catalytic activity, scanning rate, Inr strength, Pol II flux, promoter upstream and downstream constraints. See discussion in main chapter. **(D)** Increased Pol II catalytic activity will increase the TSS efficiency, therefore, usage of upstream TSSs that are first seen by Pol II will be increased, along with quickly reduced Pol II flux. The usage of downstream TSSs might be decreased due to quickly reduced Pol II flux at downstream region, resulting an upstream shifted TSS distribution as shown in green, black and green arrows. **(E)** Decreased Pol II catalytic activity will decrease the TSS efficiency, therefore, upstream TSSs will be less efficiently used and the reserved Pol II activity from failed upstream initiation will be applied to the downstream sites, apparently showing as increased TSS usage at downstream sites, but still within the window defined by TFIIH's processivity. Resulting a TSS distribution shown as blue, black and blue arrows.

### 3.1.2 Predicted roles of Ssl2 in Shooting Gallery model for TSS scanning

We hypothesize that alleles with increased Ssl2/TFIIH translocase activity could either be *scanning rate* or *scanning processivity* GOF alleles. These two types of alleles could both shift

TSS distribution downstream but through different mechanisms. Respectively, a *scanning rate* GOF allele would decrease TSS efficiency upstream by increasing the scanning speed and decreasing the exposure time of upstream template to Pol II active site. As a consequence, there will be increased TSS usage at downstream sites due to increased Pol II flux having failed to initiate upstream, similarly to Pol II LOF efficiency alleles **(Figure 22, blue TSS)**. However, a *processivity* GOF allele, which also shifts TSS distributions downstream by having increased translocase processivity, does not change the absolute efficiency of TSS usage. Instead, it expands the scanning window by allowing Pol II machinery to scan further downstream **(Figure 22, orange dotted curve)**. As a consequence, *processivity* GOF alleles increase the potential sites for scanning downstream, but can only be observed if Pol II flux (Pol II molecules still scanning) is present to support initiation **(Figure 22, orange TSS)**. Conversely, we hypothesize that alleles with decreased Ssl2/TFIIH translocase activity could either be *scanning rate* LOF alleles or *processivity* LOF alleles. Again, these two classes of alleles would both shift TSS distributions upstream but through different means and present distinct changes to TSS distributions. For example, a *scanning rate* LOF allele would slow the scanning speed and increase the exposure time of DNA template to Pol II active site, thus increasing TSS efficiency at upstream sites and causing a TSS distribution shift upstream **(Figure 22, green TSS)**. This is similar to *Pol II activity* GOF alleles' behavior in changing TSS efficiency. In contrast, a *processivity* LOF allele would limit the Pol II machinery's access to downstream TSSs sites by reducing the scanning window **(Figure 22, purple dotted curve)**. Consequently, there would be an upstream shift in TSS distribution compared to WT, without the activation of additional upstream TSSs **(Figure 22, purple TSS)**. However, there is a possibility of *scanning rate* GOF alleles not only increasing the efficiency downstream but also the distance of downstream scanning, thus making TSS window wider like *processivity* GOF

alleles. This is hypothetically achieved in *scanning rate* GOF alleles by driving Pol II further downstream through increasing the DNA translocation speed within a certain amount of time. Similar to discussions with *processivity* GOF alleles, the increased downstream TSS usage beyond normal scanning window could only be observed if there is Pol II flux going there.

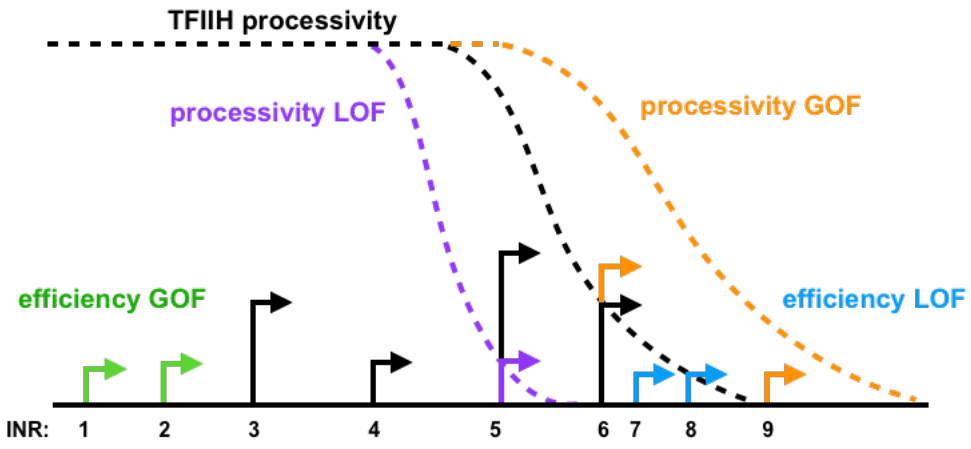



**Figure 22 Processivity and efficiency alleles shape TSS distribution differently**

Schematic showing the effect of changed processivity or efficiency on TSS distribution of a promoter. Processivity GOF alleles with increased processivity (orange dotted curve) are hypothesized to drive Pol II scanning further downstream, thus expanding the scanning window and support Pol II to use more downstream TSS (orange TSS). In contrast, processivity LOF alleles with decreased processivity (purple dotted curve) hypothetically limit Pol II scanning downstream, therefore limiting normally used downstream TSS (purple TSS). The efficiency GOF alleles can efficiently use any TSS first seen, therefore hypothetically increase TSS upstream of minor sites (green TSS). In contrast, efficiency LOF alleles are not able to efficiently use upstream TSS, unused Pol II activity applies to the downstream TSS and is observed as apparently increased downstream TSS usage (blue TSS).

### 3.1.3 *ssl2* alleles and their hypothetical functions

**(1)** In our screening, *ssl2 N230D* and similar alleles show decreased downstream TSS usage and narrower TSS windows compared to WT **(Figure 20B)**. It is likely that *ssl2 N230D* is a

*processivity* LOF allele with decreased Ssl2/TFIIH translocase activity **(Figure 22, purple curve)**. **(2)** In contrast, *ssl2 N230I* and similar alleles show increased TSS usage at downstream sites of *ADH1* promoter and shift TSS distribution downstream, behaving like a hypothetical processivity *GOF* allele **(Figure 20B)**. As discussed above, *ssl2 N230I* could also be a *scanning rate* GOF allele and show similar behavior with *Pol II* LOF alleles **(Figure 22, blue TSS)**. **(3)** Moreover, none of identified *ssl2* alleles behaves like a hypothetical *scanning rate* LOF allele as they do not increase upstream minor TSS usage at *ADH1* **(Figure 18). (4)** We observed that, compared to WT TSS usage at *ADH1* promoter, none of our *ssl2* mutants strongly used new TSSs further downstream as an apparent *processivity* GOF allele could do **(Figure 18; Figure 22, Inr 9)**. Based on what we have known from the single molecule experiment measured scanning distance distributions, there are several possibilities for this observation. i) The increased processivity is not showed as apparent new TSS sites downstream due to the distance effect. The single molecule study with yeast promoters showed that TFIIH could extend the initial transcription bubble from 24 nt to 24+96 nt long on average[44]. It was observed that TFIIH's processivity starts to decay after ~95 nt downstream scanning. If we imagine that downstream sites are at distances where TFIIH has a reduced probability of reaching, for example the downstream major and minor TSSs of *ADH1* promoter (or Inr 6-8 in Figure 22). Then increasing of TFIIH's processivity could result in increased TSS usage at such downstream sites, but not apparent new downstream site utilization because TFIIH is about to stop scanning at the downstream sites. In this case, *ssl2 N230I* could be a *processivity* GOF allele. ii) Another possibility is that *ADH1* promoter doesn't contain potential efficient Inr elements where processivity increases **(Figure 22, TSS at INR9).** iii) Alternatively, our identified alleles increase processivity, but use of new downstream TSSs doesn't appear due to lack of Pol II flux at further downstream sites. To better understand the possibilities of *ssl2*

*N230I* being either a scanning rate GOF allele or a processivity GOF allele, we designed *ssl2* genetic interaction studies and made specific predictions based on their possible roles. See discussion in 3.1.5.

### 3.1.4 Efficiency alleles and processivity alleles

In our previous studies for genetic interactions between initiation alleles, suppressive/additive interactions were broadly observed between Pol II, TFIIB and TFIIF alleles, suggesting the possibility of these factors functioning in the same pathway, hypothetically the efficiency of transcription[78]. We therefore named Pol II and TFIIB/F alleles as *efficiency* alleles. Our Shooting Gallery model predicts that activities or factors controlling scanning *processivity* would show distinct genetic effects when combined with activities or factors controlling transcription *efficiency*. Consistent with this notion and contrary to the suppressive/additive interactions between efficiency alleles, epistatic effects have been previously observed between *sub1*Δ, a deletion of the putative transcription coactivator Sub1, and *efficiency* alleles, which is explained by our hypothesis that Sub1 functions in controlling scanning *processivity* and not *efficiency*. We anticipate that hypothesized *ssl2* processivity alleles will behave like the *sub1*Δ putative processivity allele in interaction assays with *efficiency* alleles.

### 3.1.5 Design of *ssl2* genetic interaction tests and their predictions

To understand how TFIIH activity coordinates with Pol II activity and other initiation factors in determining TSS selection, we designed genetic interaction tests between *ssl2* and Pol II, TFIIB, TFIIF and *sub1*Δ alleles. As discussed above, due to distinct behavior of hypothetical

*ssl2* processivity/rate alleles and potential Pol II/GTFs efficiency alleles, we anticipated stereotypically altered TSS usage in efficiency and processivity double mutants compared to individual ones. **(1)** Processivity alleles potentially control the width of the scanning window **(Figure 23A)**, the increased processivity (processivity GOF) expands the scanning window and hypothetically increases TSS usage at downstream sites, *ssl2* N230I is such a candidate **(Figure 23B)**. In contrast, decreased processivity (processivity LOF) reduces the scanning window and is hypothesized to decrease TSS usage at downstream sites, *ssl2* N230D is likely this type **(Figure 23C)**. **(2)** Efficiency alleles potentially shape TSS distribution by changing the efficiency of TSS within TFIIH's processivity-defined scanning window, increased efficiency activates TSS usage upstream and shifts efficiency distribution upstream, like Pol II GOF alleles **(Figure 23D)**. In contrast, decreased efficiency increases apparent TSS usage downstream and shifts and flattens TSS distribution downstream, for example Pol II LOF alleles **(Figure 23E)**. **(3)** When a processivity GOF allele is combined with an efficiency GOF allele, the double mutant is expected to show processivity GOF allele's phenotype and increase TSS usage upstream, as processivity GOF alleles effect on increasing scanning window downstream will not affect upstream TSS activation **(Figure 23F)**. **(4)** When a processivity GOF allele is combined with an efficiency LOF allele, double mutant is hypothesized to show efficiency LOF allele's phenotype and increase TSS usage downstream, because processivity GOF alleles effect on increasing scanning window further downstream doesn't affect Pol II LOF allele's effect on increase TSS usage at the normal window, or even increase TSS usage in the increased processivity expanded window **(Figure 23G)**. **(5)** Combination of a processivity LOF allele with an efficiency GOF allele is expected to result in a double mutant that shows efficiency GOF allele's effect on activation upstream TSS, because processivity LOF allele's effect on limiting scanning downstream may not affect upstream TSS

usage **(Figure 23H)**. **(6)** In contrast, combination of a processivity LOF allele with an efficiency LOF allele is expected to result in a double mutant that shows processivity LOF allele's effect on truncating downstream TSS usage, due to processivity LOF allele's effect on limiting Pol II machinery going downstream **(Figure 23I)**. Alternatively, downstream shift caused by efficiency LOF allele will still happen but in a much narrower window resulting from reduced processivity of Ssl2 LOF allele. Therefore, possible additive effect is expected between downstream shifting efficiency LOF allele and upstream shifting processivity LOF allele, but not the same as would be observed from adding two efficiency alleles of opposite polarity together.

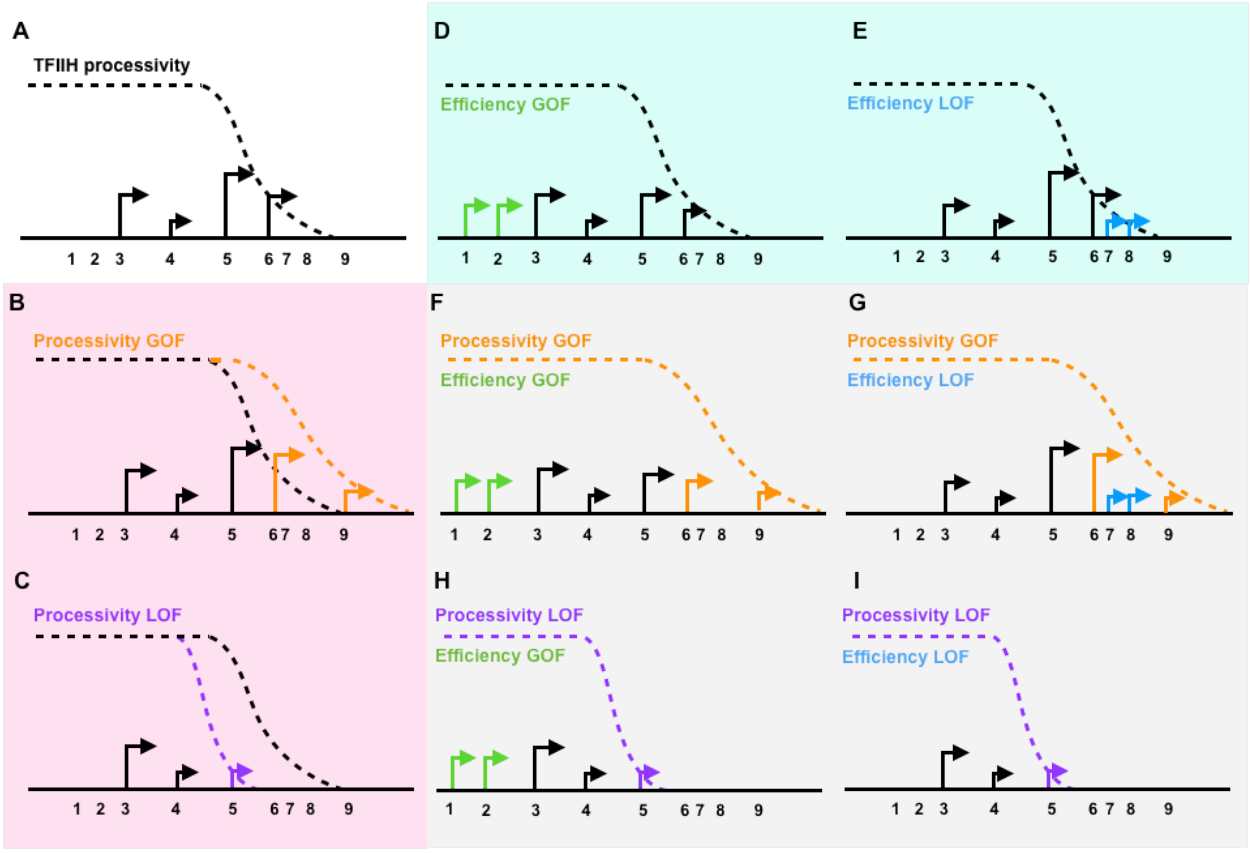



**Figure 23 Hypothesized steretypical TSS usage in processvity and efficiency double mutants**

**(A)** Schematic showing of a promoter window with TSS. Dotted curve is used to mimic WT processivity defined scanning region. **(B)** Increased processivity is hypothesized to increase TSS usage downstream. **(C)** Decreased

processivity is hypothesized to reduce TSS usage downstream. **(D)** Increased efficiency is expected to activate TSS usage upstream. **(E)** Decreased efficiency is expected to increase apparent TSS usage downstream. **(F)** Processivity GOF allele and efficiency GOF double mutant is expected to show efficiency GOF single allele's effect on activating upstream TSS. **(G)** Processivity GOF and efficiency LOF double mutant is expected to show efficiency LOF single allele's effect on increasing apparent TSS usage downstream. **(H)** Processivity LOF and efficiency GOF double mutant is expected to show efficiency GOF single allele's effect on activation TSS usage upstream. **(I)** Processivity LOF and efficiency LOF double mutant is expected to show processivity LOF single allele's effect on reducing downstream TSS usage.

## 3.2 MATERIALS AND METHODS

### 3.2.1 Yeast strains

Yeast strains used in this chapter are listed in **Table 9**.

**Table 9 Yeast strains used in Chapter III**

| Strain number | Plamids | | Genotype |
|---|---|---|---|
| CKY283 | pRP112 *RPB1* CEN *URA3* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY285 | pRP112 *RPB1* CEN *URA3* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ55 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY763 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY764 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2545 | pRP112 *RPB1* CEN *URA3* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2546 | pRP112 *RPB1* CEN *URA3* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |

| | | | |
|---|---|---|---|
| CKY2547 | pRP112 *RPB1* CEN *URA3* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2548 | pRP112 *RPB1* CEN *URA3* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2549 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2550 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2551 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2552 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2611 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY2612 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY2613 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY2614 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY2627 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY2628 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY2629 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY2630 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY2643 | pRP112 *URA3* CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY2644 | pRP112 URA3 CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |

| | | | |
|---|---|---|---|
| CKY2834 | pRP112 *URA3* CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY2835 | pRP112 *URA3* CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY2836 | pRP112 *URA3* CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY2837 | pRP112 *URA3* CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY2840 | pRP112 *URA3* CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY2841 | pRP112 *URA3* CEN ARS *RPB1* | pCK250 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY2950 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2951 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2958 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2959 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2960 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2961 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2962 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2963 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2964 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2965 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |

| | | | |
|---|---|---|---|
| CKY2968 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2969 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2970 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2971 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2972 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2973 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2974 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2975 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1* |
| CKY2982 | *RPB1* CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2983 | *RPB1* CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2990 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2991 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2992 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2993 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2994 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2995 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |

| | | | |
|---|---|---|---|
| CKY2996 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY2997 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3000 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3001 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3002 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3003 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3004 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3005 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3006 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3007 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* |
| CKY3014 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3015 | RPB1 CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3020 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3021 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3022 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3023 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |

| | | | |
|---|---|---|---|
| CKY3024 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3025 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3028 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3029 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3030 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3031 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3032 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3033 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3034 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3035 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230I* |
| CKY3042 | *RPB1* CEN LEU2 T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3043 | *RPB1* CEN LEU2 T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3048 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3049 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3050 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3051 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |

| | | | |
|---|---|---|---|
| CKY3052 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3053 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3056 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3057 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3058 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3059 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3060 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3061 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3062 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3063 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230I* |
| CKY3096 | *RPB1* CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3097 | *RPB1* CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3102 | pRS315H3alt *rpb1**XmaI Q1078S T69 corrected | pCK863 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3103 | pRS315H3alt *rpb1**XmaI Q1078S T69 corrected | pCK863 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3104 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3105 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |

| | | | |
|---|---|---|---|
| CKY3162 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3163 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3164 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3165 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3106 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3107 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3110 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3111 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3112 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3113 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3114 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3115 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3116 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3117 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 ssl2 N230D* |
| CKY3124 | *RPB1* CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3125 | *RPB1* CEN *LEU2* T69 corrected | pCK859 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |

| | | | |
|---|---|---|---|
| CKY3130 | pRS315H3alt *rpb1**XmaI Q1078S T69 corrected | pCK863 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3131 | pRS315H3alt *rpb1**XmaI Q1078S T69 corrected | pCK863 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3132 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3133 | pRS315H3alt-*rpb1** *rpb1* 11-207 (L1101S) T69 corrected | pCK864 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3166 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3167 | pRS315H3alt-*RPB1** *rpb1* 10-110 (G1097D) T69 corrected | pCK867 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3168 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3169 | pRS315H3alt-*rpb1**XmaI H1085Y T69 corrected | pCK870 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3134 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3135 | pRS315H3alt-*rpb1** 10-88 (F1086S) T69 corrected | pCK871 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3138 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3139 | pRS315H3alt-*RPB1** XmaI 1122-1123 H1085Q T69 corrected | pCK887 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3140 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3141 | pRS315H3alt-*RPB1** XmaI 1122-1123 F1084I corrected | pCK955 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3142 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3143 | pRS315H3alt-*rpb1** E1103G T69 corrected | pCK960 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |

| | | | |
|---|---|---|---|
| CKY3144 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |
| CKY3145 | pRS315H3alt-*rpb1** XmaI 1122-1123 N1082S corrected T69 corrected USE | pCK1340 | *MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128∂ gal10Δ56 rpb1Δ::CLONATMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3 ssl2 N230D* |

### 3.2.2 Yeast two-step integration

To test genetic interactions between *ssl2* allele and *rpb1*, *sua7*, *tfg2* or *tfb3* alleles, the yeast two-step integration method was used to create double mutants as previously applied in[78]. For double mutants of *ssl2* and *rpb1*, fragments containing mutations of *ssl2* N230I or N230D were cloned into vector pRS306[110] and integrated into the yeast genome through homologous recombination in a strain allowing plasmid shuffling of *rpb1* alleles. The *URA3* maker on pRS306 enables selection of successful integration by picking yeast cells with an $Ura^+$ phenotype. 5FOA was then applied on the media to select yeast cells that had lost the *URA3* marker and a copy of *SSL2* gene on the genome, either WT or mutagenized one. Loss of *URA3* marker was monitored by checking $Ura^-$ phenotypes through replica plating on yeast media containing 5FOA. Loss of a copy of *SSL2* gene was detected by genomic PCR and sequencing on the mutation site to verify retention of *ssl2* alleles. Due to the required usage of $Ura^{+/-}$ phenotypes during integration of *ssl2* mutation, a version of the *rpb1* shuffle strain that has *RPB1* cloned on pRS315 *LEU2* plasmid and the genomic copy deleted was used **(Figure 24, step 1)**. After integration of *ssl2* alleles into the genome, the pRS315 *LEU2 RPB1* plasmid was replaced with pRSII315 *URA3 RPB1*, which enables shuffling of *rpb1* alleles into the *ssl2* mutation integrated strains as described in 2.2.3 **(Figure 24, step 2)**. Two-step integration results in a double mutant strain with chromosomal *SSL2*

replaced by a *ssl2* mutant, meanwhile, chromosomal *RPB1* is deleted and the mutant form of *rpb1* is contained on the plasmid. The rationale of yeast two-step integration was used to create double mutants of *ssl2* and other alleles. Similarly, for double mutant of *ssl2* and *tfb3*, *ssl2* allele was integrated into the genome and *tfb3* allele was on plasmid. For double mutant *ssl2* and *sua7* or *tfg2*, instead, a *sua7* allele was integrated into the genome and *ssl2* alleles were introduced by plasmid shuffling.

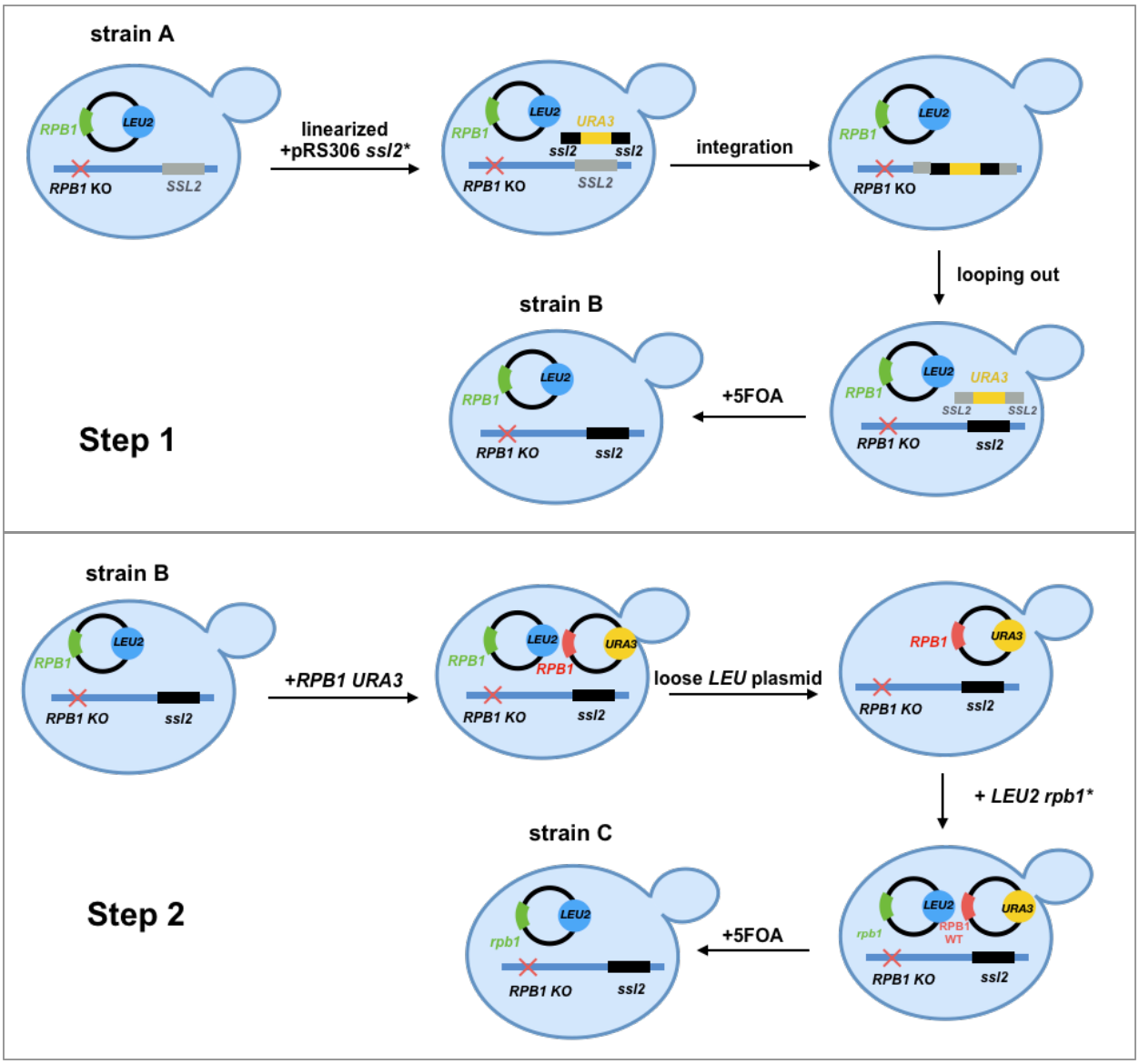



**Figure 24 Schematic illustration of yeast two step integration**

Step 1: The shuffle strain A that has *RPB1* cloned on pRS315 *LEU2* plasmid and the genomic copy deleted was used to integrate *ssl2* allele through linearized pRS306 *URA3 ssl2* plasmid guided homologous recombination. After integration, there are two copies of *SSL2* gene in yeast strain, an original WT copy and an newly integrated mutant *ssl2* allele, and the *URA3* marker. 5FOA was then applied to the medium to select yeast strains that have a copy of *SSL2* gene and the *URA3* marker lost through the Ura$^-$ phenotype. This first step results in strain B, which has genomic *RPB1* and *SSL2* deleted, a pRS315 *LEU2 RPB1* plasmid and a mutaganized *ssl2* integrated into the original *SSL2* genomic locus. Step 2: pRS315 *LEU2 RPB1* plasmid in strain B is replaced by pRS315 *URA3 RPB1* through plasmid shuffing assay. An additional plasmid shuffling assay is then performed to replace pRS315 *URA3 RPB1* with pRS315 *URA3 rpb1*. After step two, a yeast double mutant strain C is generated, which has genomic WT *SSL2* replaced with *ssl2* allele, the genomic *RPB1* deleted and the wanted *rpb1* allele on the plasmind.

## 3.3 RESULTS

### 3.3.1 Decreased TFIIH processivity is epistatic to decreased Pol II activity for downstream TSS scanning

To test genetic interactions between *ssl2* alleles and Pol II alleles, we created *ssl2* and *rpb1* double mutants by using the yeast two-step integration method as described in Materials and Methods. The transcription-related phenotypes of these *ssl2* mutants were detected using *IMD2*-based and other genetic reporters, and the TSS usage of these mutants at *ADH1* promoter were tested by primer extension, as discussed in Chapter II. First of all, comparison of the growth phenotypes between single and double *ssl2 rpb1* mutants indicates no strong genetic interactions between *ssl2* and *rpb1* alleles **(Figure 25)**. Among two tested *ssl2* alleles (N230D and N230I) and eight *rpb1* alleles (H1085Y, N1082S, H1085Q, F1086S, F1084I, L1101S, E1103G and G1097D), only one combination showed lethal phenotype, which is *ssl2* N230I and *rpb1* H1085Y. This

observation is different from our lab's previous studies, in which lethality interactions were observed between the hypothetical efficiency alleles, *e.g.* Pol II alleles and TFIIB/TFIIF alleles.

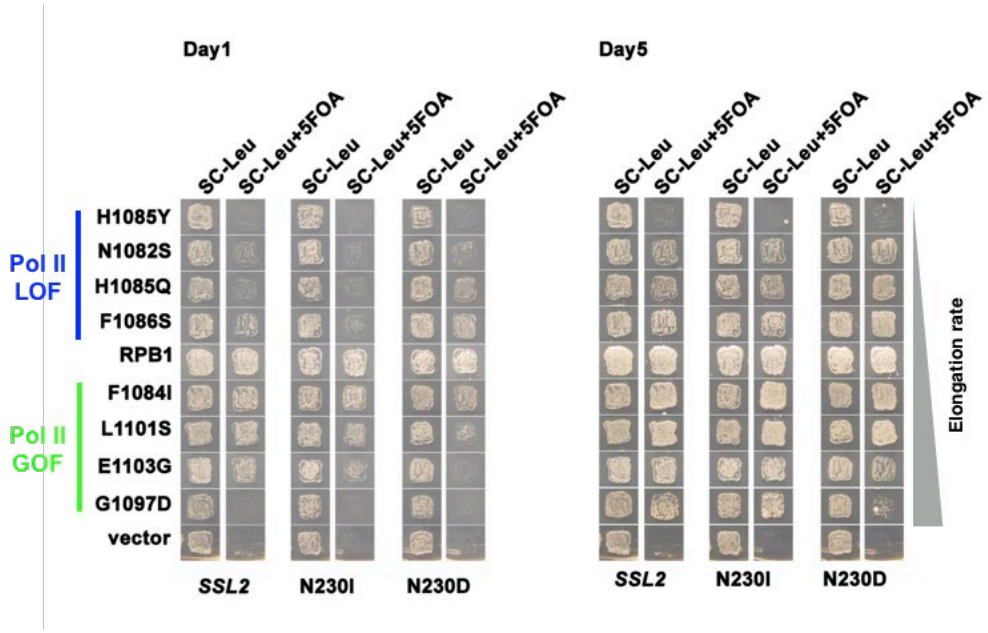



**Figure 25 Synthetic lethality was rarely observed between the interactions of *ssl2* and *rpb1* alleles**

Lethality phenotype is not extensively observed between *ssl2* and *rpb1* alleles. Patch assay shows that among all tested *ssl2* and *rpb1* double mutants, which combines *ssl2* allele of N230D or N230I, and *rpb1* allele of H1085Y, N1082S, H1085Q, F1086S, F1084I, L1101S, E1103G or G1097D, only *ssl2* N230I and *rpb1* H1085Y double mutant shows a lethal phenotype.

As hypothesized in the Shooting Gallery model, *ssl2 processivity* alleles affect the width of scanning window through increasing or decreasing the scanning distances and are genetically hypothesized to be epistatic to any alleles that change individual TSS *efficiency*, for example Pol II activity alleles. **(1) *ssl2* up + Pol II LOF = epistasis.** Consistent with this prediction, upstream shifting *ssl2* alleles hypothetically defective in processivity are found epistatic to Pol II activity LOF alleles. We observed that double mutants of a *ssl2* upstream allele and Pol II LOF alleles (*ssl2 N230D* + Pol II LOF) show MPA-sensitive phenotypes similar to the *ssl2* upstream single

mutant **(Figure 26A, purple box)**. Comparison of TSS distributions at *ADH1* between single and double mutants indicates that Pol II LOF's activity in increasing the efficiency of downstream TSSs is almost completely diminished by the putative reduced processivity of the *ssl2* upstream allele **(Figure 26B, compare Pol II LOF on left and right panel)**. This diminished downstream usage in Pol II LOF is quantitively shown as, for example in N1082S, 12% increase to 7% decrease of relative TSS usage at bin 5 of the *ADH1* promoter before and after the incorporation of *ssl2* upstream allele **(Figure 26D-G)**. **(2) *ssl2* up + Pol II GOF = epistasis/non-additivity.** Additional epistatic effects were seen between a *ssl2* upstream shifting allele and Pol II activity GOF alleles (*ssl2 N230D* + Pol II GOF). Although both single and double mutants of *ssl2* upstream allele and Pol II GOF alleles show $MPA^S$ phenotypes, we noticed that double mutants' $MPA^S$ phenotypes were not much worse than they were in Pol II GOF single mutants **(Figure 26A, green box)**. There is slight additivity of $MPA^S$ phenotypes shown in double mutants compared to *rpb1* single ones, however, this tiny additivity is different from the additivity our lab previously observed between potential efficiency alleles that lead to lethality phenotypes. Examination of TSS distribution showed that the increased efficiency at upstream sites of *ADH1* in Pol II GOF alleles is essentially unaffected when an *ssl2* upstream allele is incorporated, in spite of a dramatic decrease at downstream sites resulting from proposed reduced processivity in the *ssl2* allele **(Figure 26B, compare Pol II GOF in left and right panel, 26D-G)**. The above two interactions between putative *ssl2* processivity and Pol II activity alleles support the idea that Pol II activity is able to change TSS efficiency and can only do so within the scanning window defined by Ssl2's processivity.

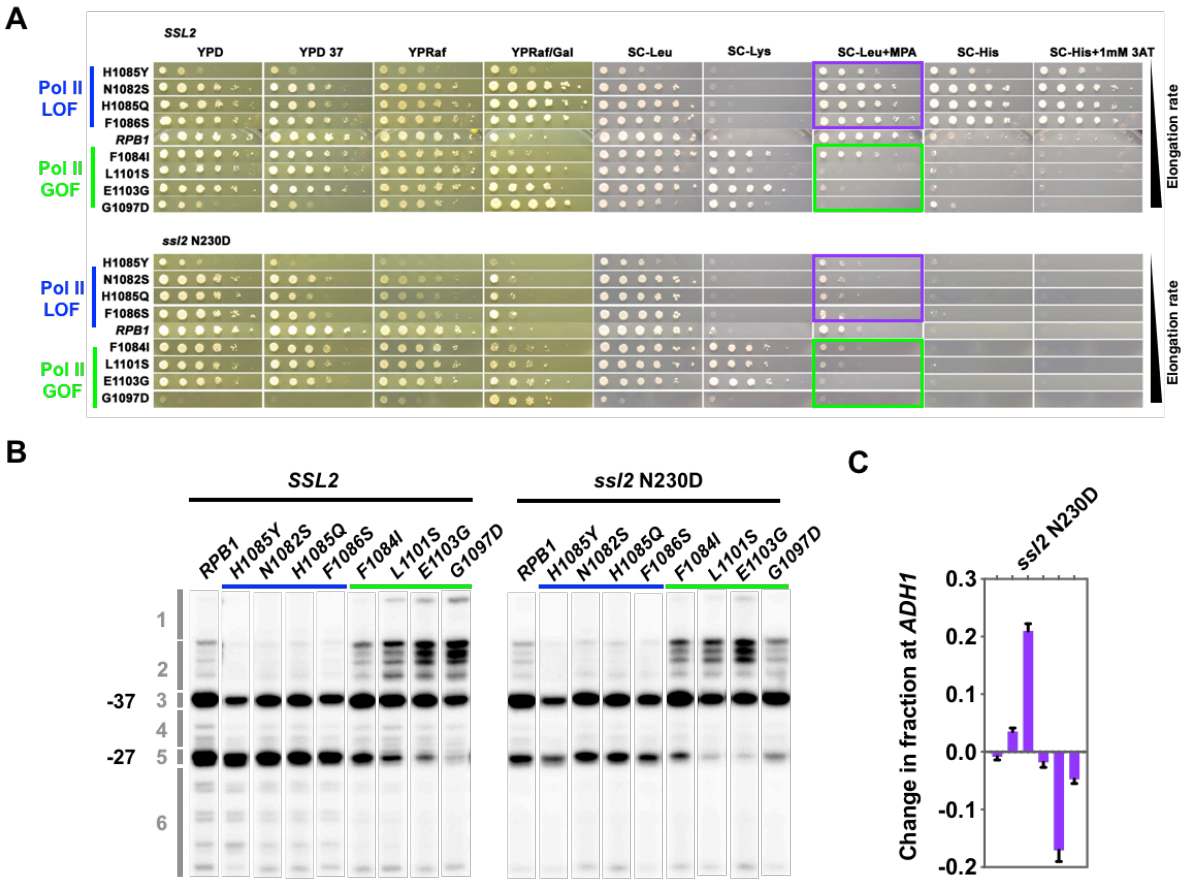

A
SSL2
YPD
YPD 37
YPRaf
YPRaf/Gal
SC-Leu
SC-Lys
SC-Leu+MPA
SC-His
SC-His+1mM 3AT
Pol II LOF
H1085Y
N1082S
H1085Q
F1086S
RPB1
Pol II GOF
F1084I
L1101S
E1103G
G1097D
Elongation rate
ssl2 N230D
B
SSL2
ssl2 N230D
RPB1
H1085Y
N1082S
H1085Q
F1086S
F1084I
L1101S
E1103G
G1097D
-37
-27
C
ssl2 N230D
Change in fraction at ADH1
D
SSL2 vs. WT
E
ssl2 N230D vs. WT
F
ssl2 N230D vs. ssl2
G
ssl2 N230D vs. rpb1

**Figure 26 Reduced TFIIH processivity limits Pol II scanning downstream**

**(A)** Ssl2 processivity LOF allele's $MPA^S$ phenotype is epistatic to Pol II LOF alleles' $His^+$ phenotypes. Spot assay shows that the hypothetical processivity LOF *ssl2* N230D allele's $MPA^S$ phenotype is retained in *ssl2* Pol II GOF double mutants, whereas Pol II GOF alleles' $His^+$ phenotypes are gone. Additionally, epistasis/non-additive interactions are observed between *ssl2* processivity LOF allele and Pol II LOF alleles. Combinations of *ssl2* processivity LOF allele ($MPA^S$) and Pol II LOF alleles ($MPA^S$) don't lead to lethal phenotypes (non-additivity), instead, double mutants show $MPA^S$ phenotypes in strength that are more similar to *ssl2* processivity LOF single mutant. **(B)** Reduced processivity in hypothetical *ssl2* processivity LOF allele limits Pol II LOF for using downstream TSS. Primer extension results show that the increased TSS usage at downstream sites of *ADH1* promoter in Pol II LOF alleles are completely diminished by incorporated *ssl2* processivity LOF allele N230D. In addition, reduce processivity at promoter downstream sites doesn't affect Pol II GOF alleles to activate TSS usage at upstream sites. The increased TSS usage at upstream sites from Pol II GOF alleles is retained after integration of *ssl2* processivity LOF allele that reduces scanning to downstream sites. **(C)** Quantification of TSS usage at *ADH1* in *ssl2* N230D single mutant. **(D)** Quantification of TSS usage at *ADH1* in *rpb1* single mutant. **(E)** Quantification of TSS usage at *ADH1* in *ssl2* N230D and *rpb1* double mutants. TSS usage in double mutants is compared to WT TSS usage. **(F)** Quantification of TSS usage at *ADH1* in *ssl2* N230D and *rpb1* double mutants. TSS usage in double mutants is compared to *ssl2* N230D single mutant. **(G)** Quantification of TSS usage at *ADH1* in *ssl2* N230D and Pol II double mutants. TSS usage in double mutants is compared to *rpb1* single mutant.

**(3) *ssl2* down + Pol II GOF = epistasis.** Continuing, epistasis is also observed when Pol II activity GOF alleles were combined with a *ssl2* downstream shifting allele (*ssl2 N230I* + Pol II GOF). In this case, Pol II activity GOF alleles' effect of activating normally poorly used upstream TSS was almost completely retained in double mutants, resulting in $MPA^S$ phenotypes as shown in Pol II GOF single mutants **(Figure 27A, green box; 27B, compare Pol II GOF in left and right panel; 27D-G)**. This result meets with our expectation for *ssl2 N230I* serving as a processivity GOF allele: the increased efficiency at upstream TSS region caused by Pol II alleles would be less affected by proposed changes to scanning processivity in the downstream region

caused by the *ssl2* allele. We take this observation as an evidence for *ssl2 N230I* not being a scanning *rate* allele. A presumed scanning rate allele should alter TSS efficiency allele and therefore we would expect additive effects with other efficiency alleles, but epistasis is observed here. **(4) *ssl2* down + Pol II LOF = epistasis (slight additive, not lethal).** Moreover, when downstream shifting alleles of Pol II activity LOF and *ssl2* that both show $His^+$ phenotypes are combined (*ssl2 N230I* + Pol II LOF), double mutants show only slightly stronger $His^+$ phenotypes and little stronger TSS downstream shift than the *ssl2* single mutant **(Figure 27A, blue box, 27B, compare Pol II LOF in left and right panel; 27D-G)**. These results suggest that the *ssl2* downstream allele and Pol II LOF alleles are non-additive or epistatic. This is consistent with the rationale that Pol II LOF alleles are able to increase TSS efficiency at downstream sites ($His^+$ phenotypes), but not beyond the scanning window defined by *ssl2* downstream allele's processivity (no stronger than *ssl2* downstream allele's $His^+$ phenotypes). Additionally, as we mentioned earlier, two hypothetical processivity alleles, *ssl2* and *sub1*Δ, should behave similarly in genetic interactions with efficiency alleles. Surprisingly, we noticed that the epistasis or non-additivity between *ssl2 N230I* and Pol II LOF alleles is different from the lethality phenotypes between *sub1Δ* and Pol II LOF alleles **(see discussion in 3.3.4)**[78]. This observation is not against our hypothesis for both of them being processivity alleles. Instead, *sub1Δ* by itself is such a strong downstream shifting allele and could have multiple effects, making it possible for any tiny further TSS downstream shift to become lethal. To sum up, we observe mainly epistatic effects between *ssl2* and Pol II alleles, which hypothetically control scanning processivity and TSS efficiency, respectively.

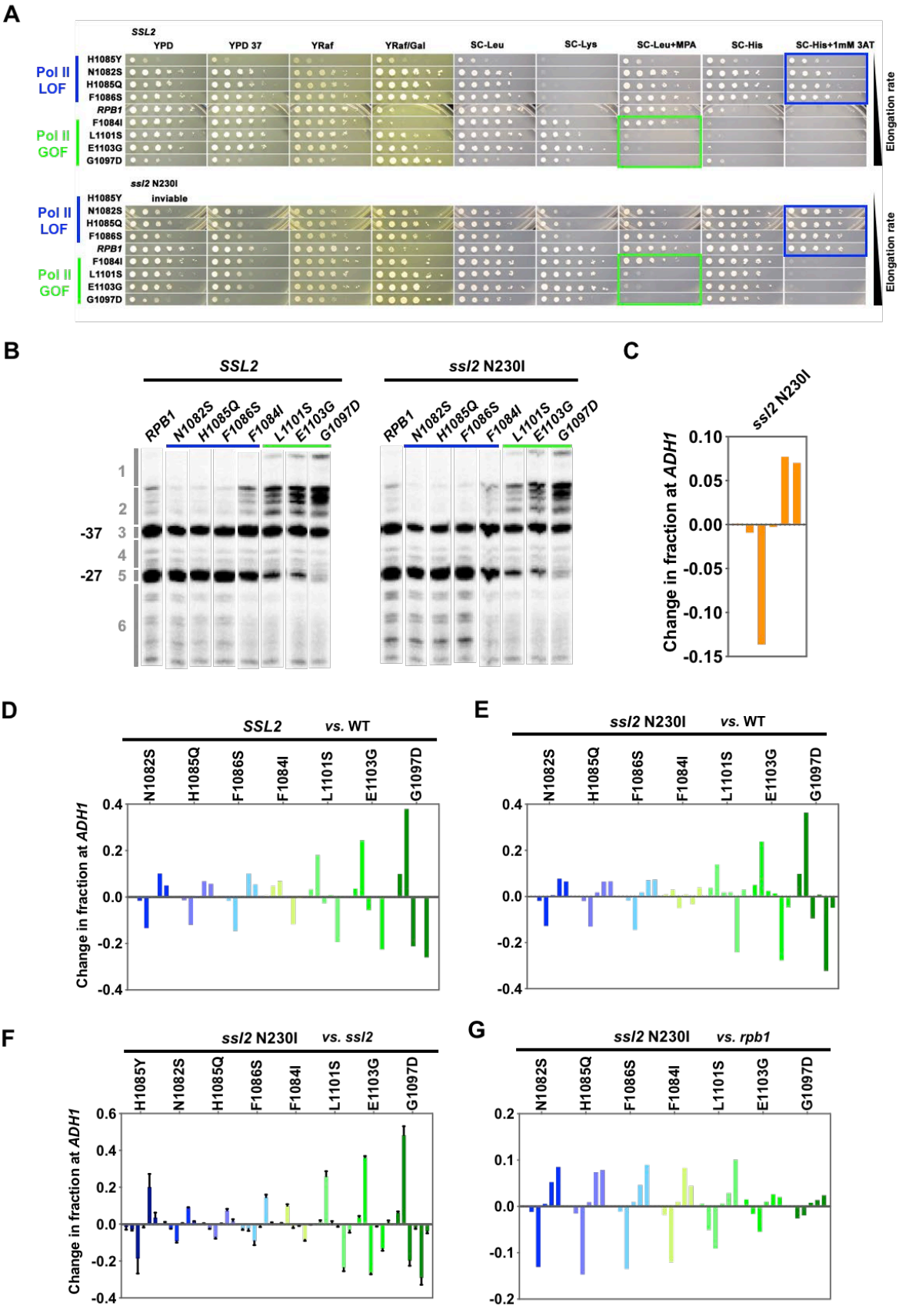

A
SSL2
YPD
YPD 37
YRaf
YRaf/Gal
SC-Leu
SC-Lys
SC-Leu+MPA
SC-His
SC-His+1mM 3AT
Pol II LOF
Pol II GOF
H1085Y
N1082S
H1085Q
F1086S
RPB1
F1084I
L1101S
E1103G
G1097D
Elongation rate
ssl2 N230I
inviable
B
SSL2
ssl2 N230I
RPB1
N1082S
H1085Q
F1086S
F1084I
L1101S
E1103G
G1097D
-37
-27
C
ssl2 N230I
Change in fraction at ADH1
D
SSL2 vs. WT
E
ssl2 N230I vs. WT
F
ssl2 N230I vs. ssl2
G
ssl2 N230I vs. rpb1

**Figure 27 Pol II efficiency alleles are able to increase TSS efficiency within the processivity defined scanning window**

**(A)** Pol II GOF alleles' $MPA^S$ phenotypes is epistatic to Ssl2 processivity GOF allele's $His^+$ phenotype. Spot assay shows that the Pol II GOF alleles' $MPA^S$ phenotypes are retained in *ssl2* Pol II GOF double mutants, whereas hypothetical processivity GOF *ssl2* N230I allele's $His^+$ phenotype is gone. Additionally, epistasis/non-additivity interactions are observed between *ssl2* processivity GOF allele and Pol II LOF alleles. Combinations of *ssl2* processivity GOF allele ($His^+$) and Pol II LOF alleles ($His^+$) don't lead to lethal phenotypes (non-additivity), instead, double mutants show $His^+$ phenotypes in strength more like processivity GOF single alleles. **(B)** Both Pol II LOF alleles' effects on increasing TSS usage at upstream and GOF alleles' effects on increasing apparent downstream TSS usage are seen when TFIIH's processivity is increased. Primer extension results show that the increased TSS usage at upstream or downstream sites of *ADH1* promoter in Pol II GOF or LOF alleles respectively, are retained by incorporated *ssl2* processivity GOF allele N230I. **(C)** Quantification of TSS usage at *ADH1* in *ssl2* N230I single mutant. Only one biological replicate in C-G is done for the purpose of preliminary test, however, the TSS usage pattern is very consistent between the same class of alleles, I'm currently repeating this experiment using three biological replicates. **(D)** Quantification of TSS usage at *ADH1* in *rpb1* single mutant. **(E)** Quantification of TSS usage at *ADH1* in *ssl2* N230I and *rpb1* double mutants. TSS usage in double mutants is compared to WT TSS usage. **(F)** Quantification of TSS usage at *ADH1* in *ssl2* N230I and *rpb1* double mutants. TSS usage in double mutants is compared to *ssl2* N230I single mutant. **(G)** Quantification of TSS usage at *ADH1* in *ssl2* N230I and Pol II double mutants. TSS usage in double mutants is compared to *rpb1* single mutant.

### 3.3.2 Lethality between upstream shifting *ssl2* alleles and TFIIB allele invokes a discussion of TFIIB function in regulating Pol II efficiency

We anticipated that *sua7-1*, encoding a mutant TFIIB, which was initially defined as an efficiency allele due to its additive behavior with Pol II efficiency alleles, would also show epistasis effects when interacting with *ssl2* processivity alleles. **(1) *sua7-1* + *ssl2* downstream = lethality.** Notably, when *sua7-1* is combined with any of our tested *ssl2* downstream shifting

alleles, which are hypothetical processivity GOF (*sua7-1* + *ssl2* I170T/V226D/N230I/E340G/L225P/R636C), lethal phenotypes are observed **(Figure 28A, red boxes)**. This lethality is different from the epistasis effect when the same *ssl2* downstream alleles are combined with Pol II efficiency alleles **(Figure 27A)**. However, lethal phenotypes were previously observed by us and others when *sua7-1* was combined with other downstream shifting alleles: Pol II LOF alleles and *sub1Δ* **(Figure 29)**[78,111]. These observations together indicate potential complexity for Sua7/TFIIB in promoter scanning and TSS selection: in addition to altering TSS usage efficiency, *sua7-1* could have additional defects rendering it sensitive to increased *ssl2* processivity, for example, it might affect PIC integrity. In this case, its additional defects might allow it to have stronger genetic interactions, which would be consistent with observed lethality between all the other downstream shifting alleles, either efficiency or processivity alleles. **(2) *sua7-1* + *ssl2* upstream = epistasis.** In our genetic test, *sua7-1* downstream shifting allele ($His^+$) was found showing epistasis effect with *ssl2* upstream shifting processivity alleles ($MPA^S$). We observed that double mutants of *sua7-1* and *ssl2* upstream alleles (*sua7-1* + *ssl2* N230D/V473D/D522V/Y750*) phenocopy *ssl2* single mutants' $MPA^S$ phenotypes **(Figure 28B, purple boxes)**. This observation is consistent with the Shooting Gallery model's prediction that if *sua7-1* was an efficiency allele, its activity in increasing downstream TSS usage should be constrained by the reduced processivity from the putative of *ssl2* processivity LOF alleles. However, despite some level of additivity rendered by *ssl2* upstream shifting alleles, double mutants show an overall downstream TSS shift at *ADH1* promoter. Therefore, *sua7-1* is epistatic to *ssl2* upstream shifting alleles at *ADH1* promoter, which is opposite to the observation that *ssl2* upstream alleles are epistatic to *sua7-1* at *IMD2* promoter **(Figure 28C, 28D)**. This is the first discrepancy we've observed between the *IMD2* growth-phenotype predicted TSS patterns and

primer extension-detected TSS patterns at *ADH1*. In previous studies, MPA$^S$ phenotype was always a predictor of upstream TSS shifts at *ADH1*, while His$^+$ phenotype was highly-correlated with downstream TSS shifts at *ADH1*. The discrepancy indicates a potential distinction between the behavior at different promoters. For example, the discrepancy between *IMD2*-related MPA$^S$ phenotype and *ADH1* promoter TSS usage could be explained by the difference of the architectures between these two promoters. Our lab's previous study found that MPA induction caused WT yeast to shift TSS at *imd2::HIS3* reporter from the TATA-proximal "G" to downstream "A" and produced a functional *HIS3* transcript that renders yeast cell His$^+$ phenotypes[102]. However, MPA induction at such *imd2::HIS3* reporter in Pol II GOF alleles only shift TSS in between of upstream "G" and downstream "A", which is not downstream enough to give yeast a His$^+$ phenotype. Here, similarly, in the presence of *ssl2* upstream alleles, *sua7-1* allele's effect on shifting TSS downstream at *ADH1* is hindered and the downstream shift is not enough to render double mutants His$^+$ phenotypes, whereas this shift is seen at *ADH1* promoter by primer extension detected TSS usage.

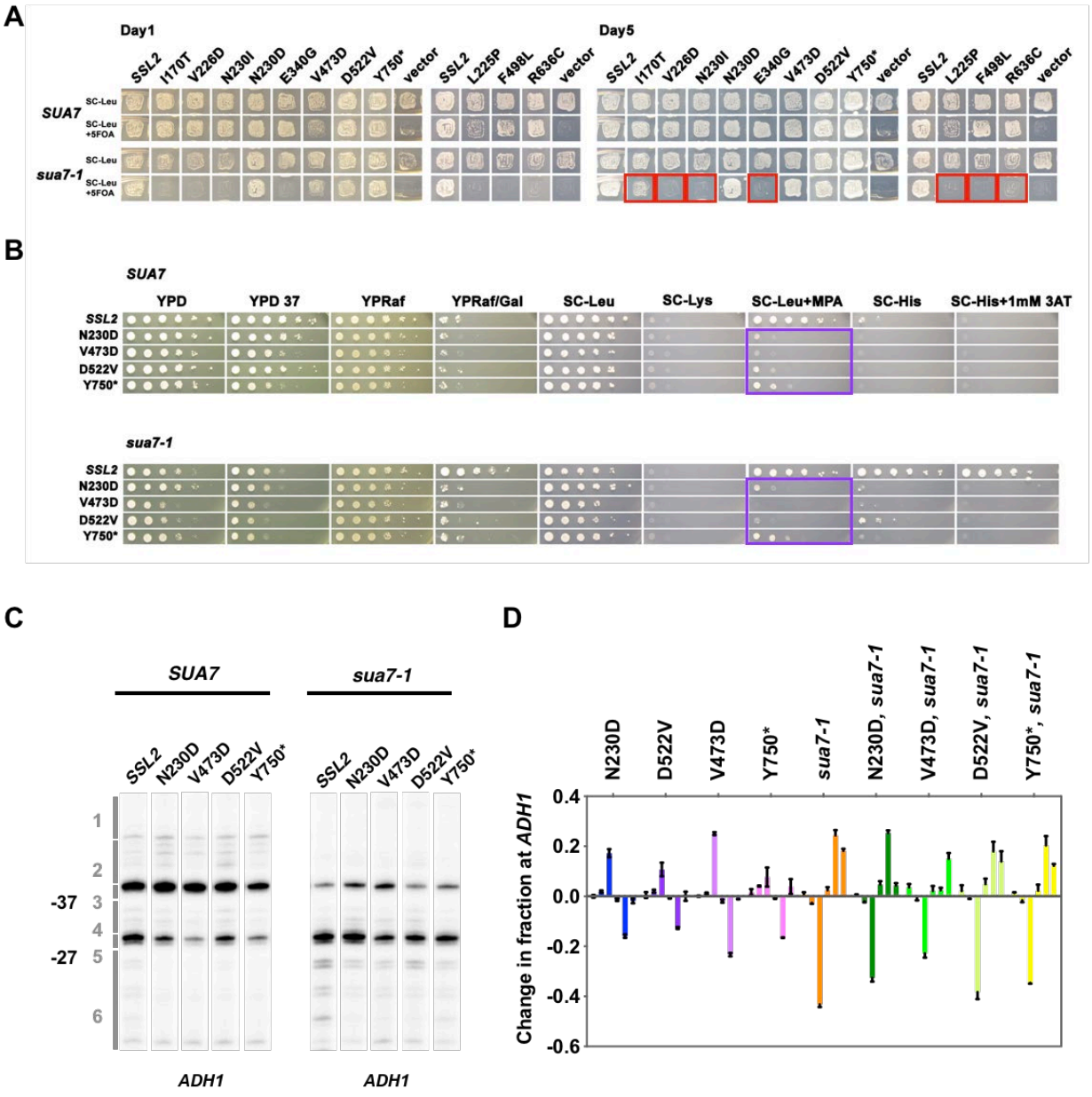



**Figure 28 The *sua7* allele shows strong and distinct genetic interaction behavior with *ssl2* alleles**

**(A)** Lethality is broadly observed between *ssl2* downstream shifting alleles and *sua7-1* downstream shifting allele. Patch assay show that the six downstream shifting alleles I170T, V226D, N230I, E340G, L225P and R636C, show synthetic lethality phenotypes when combined with *sua7-1* allele (*ssl2* I170T and *sua7-1* double mutant is very sick on 5FOA plate and become inviable when streaking on an extra 5FOA plate for single colonies). Four *ssl2* upstream shifting alleles N230D, V473D, D522V, Y750* show normal growth phenotypes when *sua7-1* is incorporated. One *ssl2* upstream shifting allele F498L show abnormal (lethality) genetic interactions compared to other *ssl2* upstream

shifting alleles (normal growth). This allele is being a unique one that also shows lethality with TFIIF and Sub1 alleles as discussed in the corresponding sections, respectively. **(B)** *ssl2* upstream shifting alleles show epistasis effect with *sua7-1* downstream shifting allele for $MAP^S$ phenotypes. Spot assay shows that four upstream shifting *ssl2* alleles' strong $MPA^S$ phenotypes (N230D, V473D, D522V and Y750*) are almost completely unaffected when *sua7-1* allele, which shows a $His^+$ phenotype by itself, is incorporated. **(C)** Double mutants of *ssl2* upstream shifting allele and *sua7-1* downstream shifting allele shift TSS distribution downstream. Primer extension results show that *ssl2* alleles' effect on shifting TSS distribution upstream is almost completely reversed to downstream shifting after incorporation of *sua7-1* allele, with only slight additive effect seen in double mutants' TSS usage. **(D)** Quantification of primer extension detected TSS usage at *ADH1* in *ssl2* N230D, *sua7-1* single or double mutants.

### 3.3.3 Distinct genetic interactions between *ssl2* and TFIIF alleles reveal a potential dual function of TFIIF in TSS selection

As discussed earlier in "Efficiency alleles and processivity alleles" (3.1.4), TFIIF is a hypothetical efficiency factor due to its alleles' additive behavior with Pol II efficiency alleles. The shooting gallery model predicts that alleles of TFIIF would show epistasis effects when interacting with *ssl2* processivity alleles. **(1) *tfg2*Δ*146-180* + *ssl2* upstream = lethality.** Strikingly, when *tfg2*Δ*146-180* is combined with *ssl2* upstream shifting alleles, which are hypothetical processivity LOF (*tfg2*Δ*146-180* + *ssl2* N230D/V473D/F498L/D522V/Y750*), lethal phenotypes are observed **(Figure 27A, red boxes)**. Although double mutant of *ssl2* Y750* and *tfg2*Δ*146-180* showed moderate level of growth in patch assay, further growth on additional yeast medium gives very tiny and sick colonies. The strong genetic interactions between *tfg2*Δ*146-180* and *ssl2* upstream shifting alleles is again different from the epistasis effect we observed from interaction of hypothetical Pol II efficiency alleles and the putative *ssl2* processivity alleles **(Figure 25A)**. We proposed two possibilities for this observation: firstly, *tfg2*Δ*146-180* could have some additional

defects causing increased sensitivity to *ssl2* upstream shifting alleles, for example, defects in PIC integrity; secondly, *tfg2*Δ*146-180* could have some processivity defects, which are different from the processivity defects from *ssl2*, working concurrently to cause lethal phenotypes. **(2) *tfg2*Δ*146-180* + *ssl2* downstream = epistasis/additivity.** We observed that double mutants of *tfg2*Δ*146-180* and *ssl2* downstream alleles (*tfg2*Δ*146-180* + *ssl2* L225P/N230I/R636C) show $MPA^S$ phenotypes that are similar to *tfg2*Δ*146-180* single mutants' $MPA^S$ phenotypes **(Figure 27B, orange boxes)**. This epistasis effect between *tfg2*Δ*146-180* and the *ssl2* downstream processivity alleles in terms of the $MPA^S$ phenotype is consistent with the predication of *tfg2*Δ*146-180* being an upstream shifting efficiency allele and less likely affected by the processivity allele's effect on TSS usage at downstream sites. However, the $MPA^S$ phenotypes observed in *tfg2*Δ*146-180* and *ssl2* downstream double mutants are not as strong as *tfg2*Δ*146-180* single mutants' $MPA^S$ phenotypes, indicating some extent of additivity conferred by *ssl2* downstream shifting alleles. This observation suggests a possibility of *tfg2*Δ*146-180* having some defects in processivity and this defect in processivity is suppressed by the potentially increased processivity of *ssl2* downstream alleles at some extent. Together, suggesting a potential dual role of TFIIF in controlling TSS usage, either as an efficiency factor or as a processivity factor.

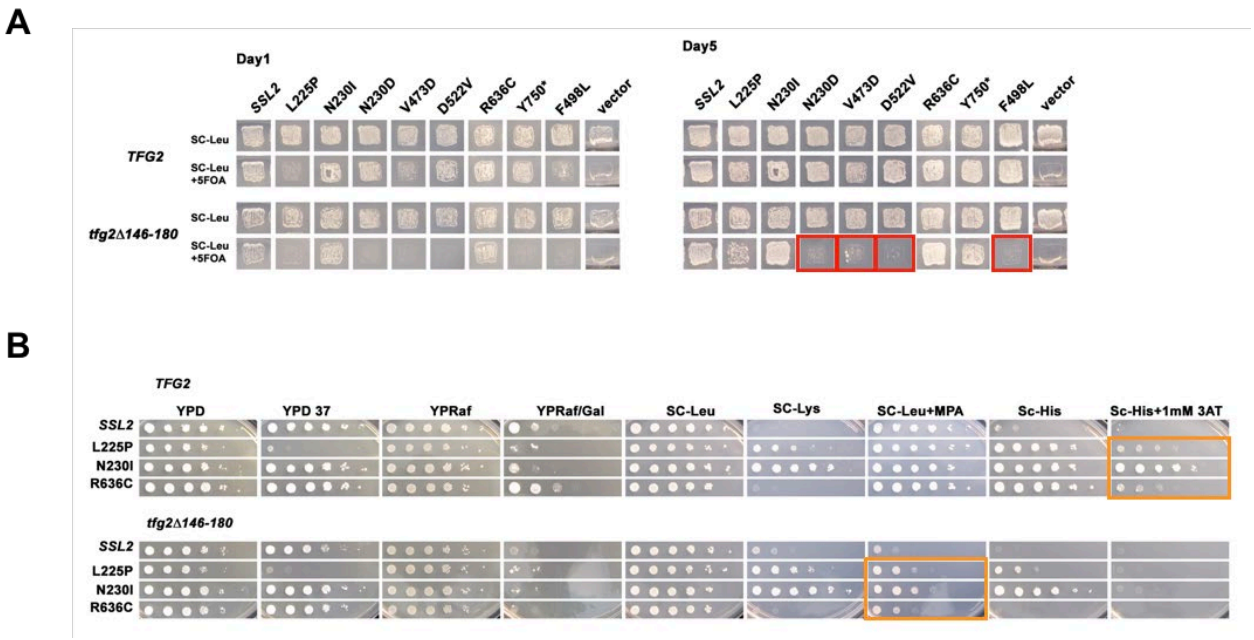



**Figure 29 Genetic interactions between *ssl2* and *tfg2* alleles reveal a potential dual function of TFIIF in TSS selection**

**(A)** Lethality is observed between *ssl2* upstream shifting alleles and *tfg2Δ146-180* upstream shifting allele. Patch assay show that upstream shifting alleles N230D, V473D, D522V and F498L show synthetic lethality phenotypes when combined with *tfg2Δ146-180* allele (*ssl2* Y750* and *tfg2Δ146-180* double mutant shows moderate level of growth on 5FOA plate, however, become very sick when streaking on an extra 5FOA plate for single colonies). Three *ssl2* downstream shifting alleles L225P, N230I and R636C show normal growth phenotypes when *tfg2Δ146-180* is incorporated. (R636C in double mutant shows slight sicker growth phenotype than single one). **(B)** *ssl2* downstream shifting alleles show both epistasis and additive effect with *tfg2Δ146-180* downstream shifting allele. Spot assay shows that the double mutants of *ssl2* downstream alleles and *tfg2Δ146-180* allele show $MAP^S$ phenotypes; however, double mutants' $MAP^S$ phenotypes are not as strong as observed in single *tfg2Δ146-180* allele.

### 3.3.4 Epistasis analysis between *ssl2* and *sub1Δ* links Sub1 to TFIIH function

*sub1Δ*, a deletion of the gene encoding Sub1 was previously found to facilitate Pol II transcription, show $His^+$ phenotype for the *imd2Δ::HIS3* fusion reporter gene, and shift TSS

distributions downstream at *ADH1*[95]. Our prior research found that when *sub1Δ* was combined with Pol II GOF alleles, epistatic effects were observed, leading to the proposal that *sub1*Δ effects in initiation were distinct from TFIIB or TFIIF alleles, as these showed additivity for initiation defects with Pol II alleles[78]. Because we have observed similar epistatic interactions between *ssl2* and Pol II efficiency alleles, we considered that Sub1 acts different from efficiency factors in controlling TSS selection, potentially behaving as a processivity factor. If true, possible additive effects could be observed between two types of processivity alleles, namely *ssl2* alleles and *sub1Δ*. However, multiple effects were observed between *ssl2* and *sub1Δ* alleles. First of all, no strong genetic interactions (lethality) were observed between *ssl2* and *sub1Δ* alleles **(Figure 28A)**. **(1) *sub1Δ* + *ssl2* up = epistasis/additivity.** In our genetic tests, upstream shifting *ssl2* alleles with $MPA^S$ phenotypes were epistatic to *sub1Δ* for its $His^+$ phenotype, showing $MPA^S$ phenotypes in double mutants (*sub1Δ* + *ssl2* N230D/V473D/D522V) **(Figure 28B, purple box)**. However, additive effects were observed in double mutants causing TSS distributions at *ADH1* to be similar to WT **(Figure 28C)**. Unlike other upstream shifting *ssl2* alleles, *sub1*Δ was epistatic to *ssl2* Y750*, meaning the double mutant (*sub1Δ* + *ssl2* Y750*) showed a $His^+$ phenotype and shifted *ADH1* TSS distribution downstream similarly to *sub1Δ* alone (only slight additivity in $His^+$ phenotype and *ADH1* TSS shifts) **(Figure 28B, C, D). (2) *sub1Δ* + *ssl2* down = epistasis/additivity.** Furthermore, epistasis was observed between *sub1Δ* and *ssl2* downstream alleles. Double mutants of *sub1Δ* and *ssl2* downstream shifting alleles (*sub1Δ* + *ssl2* L225P/N230I/R636C) show $His^+$ phenotypes similar to the *sub1Δ* single mutant **(Figure 28B, orange box)**. Slight additive effects were again observed in double mutants as TSS distribution at *ADH1* downstream moved downstream **(Figure 28C, D)**. Together, it seems that Ssl2/TFIIH controls all initiation factors for reaching downstream, including Sub1, demonstrated by *ssl2*

alleles' MPA$^S$ phenotypes and the upstream shifting effect being epistatic to *sub1Δ* allele's His$^+$ phenotype and the downstream shifting effect. Once Sub1 reaches downstream sites, it has the potential to expand scanning window further, suggested by the stronger His$^+$ phenotype and the further downstream shifting added to the *ssl2* downstream shifting alleles when *sub1Δ* is incorporated. Finally, we noticed that, different from additive/suppressive interactions between efficiency alleles, epistasis effects are mainly observed among processivity alleles.

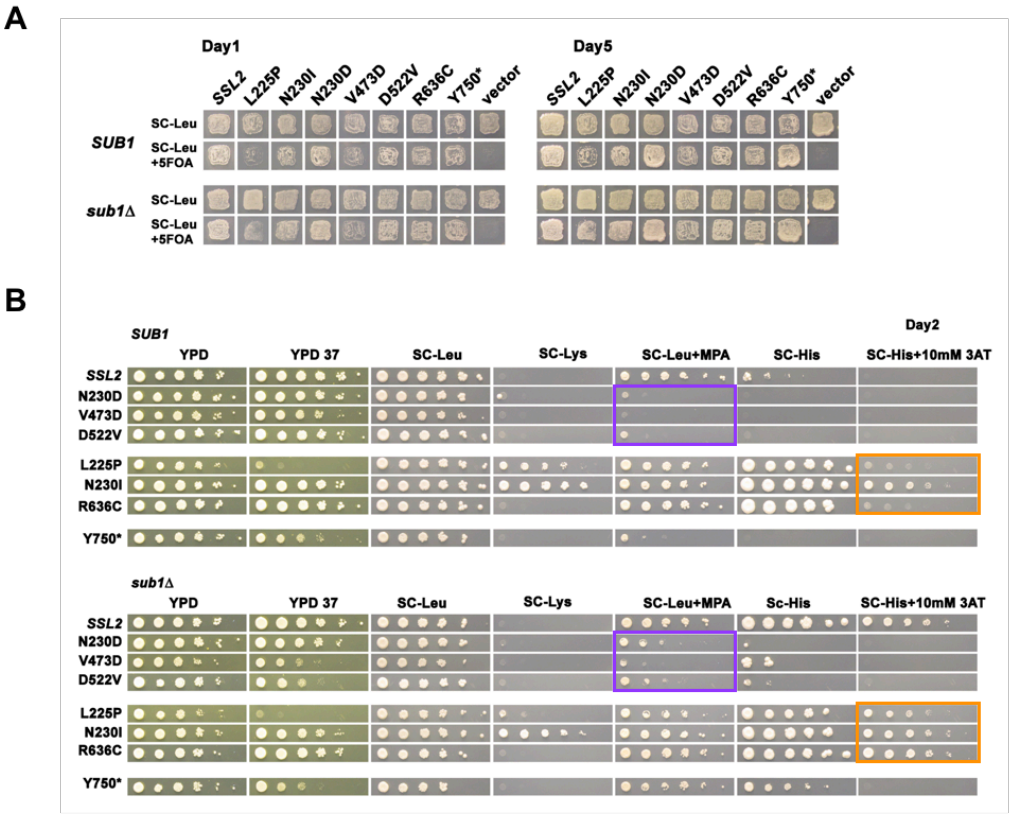

A
Day1
Day5
SSL2
L225P
N230I
N230D
V473D
D522V
R636C
Y750*
vector
SUB1
sub1Δ
SC-Leu
SC-Leu +5FOA
B
SUB1
sub1Δ
YPD
YPD 37
SC-Leu
SC-Lys
SC-Leu+MPA
SC-His
Sc-His
Day2
SC-His+10mM 3AT
SSL2
N230D
V473D
D522V
L225P
N230I
R636C
Y750*
C
SUB1
sub1Δ
-37
-27
1
2
3
4
5
6
ADH1
D
Change in fraction at ADH1
sub1Δ
N230D
V473D
D522V
N230D, sub1Δ
V473D, sub1Δ
D522V, sub1Δ
L225P
N230I
R636C
L225P, sub1Δ
N230I, sub1Δ
R636C, sub1Δ
Y750*
Y750*, sub1Δ
0.2
0.0
-0.2
-0.4

**Figure 30 Multiple genetic interactions between *ssl2* and *sub1Δ* alleles**

**(A)** No strong genetic interactions (lethality phenotype) are observed between *ssl2* and *sub1Δ* alleles. Patch assay show that double mutants of *sub1Δ* and *ssl2* L225P, N230D, N230I, V473D, D522V, R636C or Y750* show normal growth phenotypes compared to the correspondent *ssl2* single mutants. **(B)** Three types of genetic interactions are observed between *ssl2* and *sub1Δ* alleles in *IMD2* promoter. First, three *ssl2* upstream shifting alleles (N230D, V473D or D522V) are epistatic to *sub1Δ* allele in genetic phenotypes. Spot assay shows that the double mutants of *sub1Δ* allele ($His^+$) and *ssl2* upstream shifting alleles ($MPA^S$) show $MPA^S$ phenotypes like *ssl2* upstream single alleles. Second, *sub1Δ* allele is epistatic to *ssl2* allele Y750*. Double mutant of *sub1Δ* ($His^+$) and Y750* ($MPA^S$) shows a $His^+$ phenotype. Third, *sub1Δ* allele is epistatic to *ssl2* downstream shifting alleles (L225P, N230I, R636C). Double mutants of *sub1Δ* ($His^+$) and *ssl2* downstream shifting alleles ($His^+$) show $His^+$ phenotypes that are not stronger than in single *sub1Δ* allele. **(C)** Genetic interactions between *ssl2* and *sub1Δ* alleles at *ADH1* promoter. First, additive effects are observed between *ssl2* upstream shifting alleles and *sub1Δ* allele. Double mutants of *ssl2* upstream shifting alleles and *sub1Δ* downstream alleles show a compromised TSS distribution. Second, additive effects are also observed between *ssl2* downstream shifting alleles and *sub1Δ* allele. Double mutants show stronger downstream shift than either single ones. Third, epistasis effect is observed between *ssl2* Y750* and *sub1Δ.* In the double mutant, *ssl2* Y750* allele's effect on shifting TSS distribution upstream is completely reversed to downstream shifting by *sub1Δ*. **(D)** Quantification of primer extension detected TSS usage at *ADH1* in *ssl2*, *sub1Δ* single or double mutants.

## 3.4 DISCUSSION

In our genetic interaction tests between initiation alleles, we previously observed mainly suppressive/additive interactions between the hypothetical efficiency alleles: Pol II, TFIIB and TFIIF **(Figure 31)**. Here, in genetic interaction tests between *ssl2* and other initiation factors, epistatic interactions were observed among potential *ssl2* processivity alleles and *sub1*Δ. In addition, between efficiency and processivity alleles, epistatic interactions were primarily

observed. Together, our genetic interaction data between initiation factors reveals two major networks controlling promoter scanning and TSS selection: one controlling the efficiency of initiation through Pol II activity or factors regulating Pol II's activity; another network controls the processivity of initiation by TFIIH. As discussed in the genetic interactions between *ssl2* allele and *tfg2* allele, TFIIF could have dual function in TSS selection, either as an efficiency or processivity allele (which is not shown in below figure).

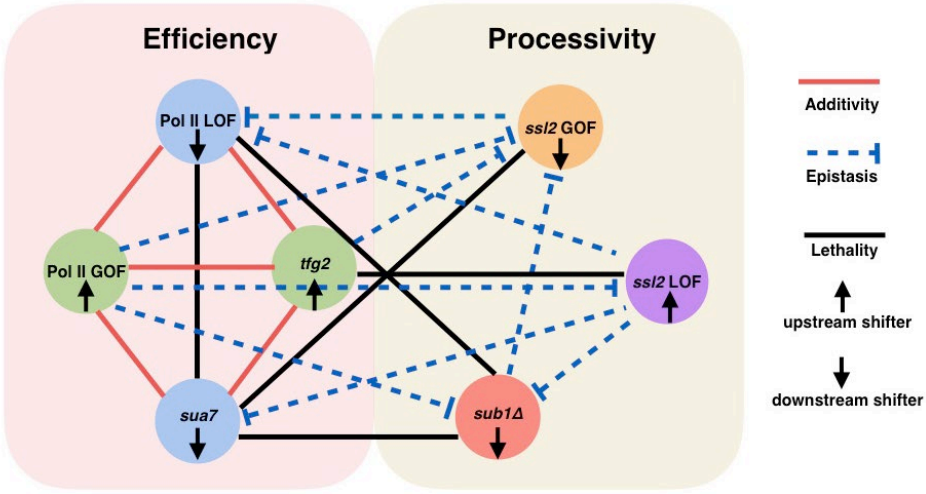



**Figure 31 Two major protein networks in controlling TSS selection**

In our genetic experiments, alleles regulating Pol II efficiency and alleles affecting TFIIH processivity are shown in two major panels with different colors, pink and beige. Additivity/suppression effects are mainly observed between the hypothetical efficiency alleles, whereas epistasis effects are mainly observed among processivity alleles or between efficiency and processivity alleles.

# 4.0 GENOME WIDE IMPACT OF SSL2 MUTANTS ON TRANSCRIPTION START SITE USAGE

## 4.1 INTRODUCTION

We have defined two potential classes of *ssl2* alleles in terms of their functions in Pol II scanning and propose they are TFIIH processivity alleles that control TSS distributions by narrowing or widening scanning windows at promoters. Representatives of these two classes and their TSS effects have been tested at the model genes of *IMD2* and *ADH1* in *S. cerevisiae*. To gain an insight into the impact of TFIIH's activity on TSS usage genome-wide, we have examined 5' ends of RNA transcripts for these two classes of *ssl2* allele in *S. cerevisiae*. We first investigated TSS distribution in two classes of *ssl2* alleles by examining the reads distribution at 5979 yeast promoters **(Figure 32A)**. We also compared the shift of TSS distribution between WT and *ssl2* mutants. As observed in model gene promoters, *ssl2* alleles that shift TSS distribution upstream or downstream at *IMD2* or *ADH1* promoters also shift TSS distribution at most of the other examined gene promoters. We therefore measured the degree of TSS shift in *ssl2* alleles by comparing the median TSS positions of the mutant and the WT **(Figure 32B)**. We also asked if the degree of TSS shift correlated with the architecture of the promoter, for example, the distance between the PIC and the dominant TSS (PIC-TSS distance), so we also examined the relationship between the degree of the TSS shift and the distance between PIC-TSS distance **(Figure 32C)**. Our Shooting Gallery model predicts that the hypothetical *ssl2* processivity GOF alleles have the potential to expand the scanning window and the *ssl2* processivity LOF alleles would reduce the scanning window. To test these predictions, we examined the width of the scanning window by measuring

the distribution of 80% TSS spanning region in candidate *ssl2* processivity alleles genome-wide **(Figure 32D)**.

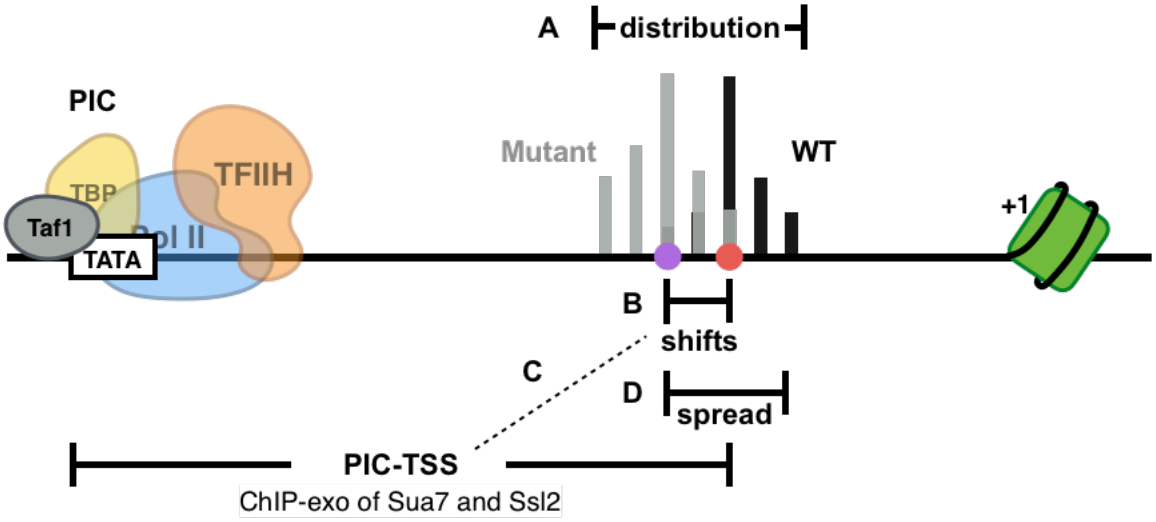



**Figure 32 Schematic showing of a yeast promoter and the TSS-seq data analysis done in this chapter**

**(A)** TSS distribution in both WT and *ssl2* mutants are examined by measuring the mapped 5' end distribution at 5979 yeast gene promoters. The shift of TSS distribution is also compared between WT and *ssl2* mutants by comparing the reads aboundance at each nucleotide position of every 5979 promoter window. (**B**) Shift of the TSS distribution between the *ssl2* mutants and the WT is measured by comparing the median TSS posistions in two strains genome-wide. **(C)** The relationship between the degree of TSS shift and the PIC-TSS distance is exmined in *ssl2* mutants. **(D)** The scanning window is measured in *ssl2* mutants by examining the genome location of the majority (80%) of TSS locations.

## 4.2 MATERIALS AND METHODS

### 4.2.1 Transcription start site sequencing (TSS-seq)

Yeast cell cultures were grown in triplicates and cells were harvested at mid-log phase at a

density of $1\times10^7$ cells/mL, as determined by cell counting. For *S. cerevisiae* TSS-seq, cells collected from 50 mL of *S. cerevisiae* culture and 5 mL of *S. pombe* culture were mixed and total RNA was extracted as described[112]. We performed cDNA library construction for TSS-seq essentially as described by Irina et.al[113], steps are described as follows. 100 ug of the isolated total RNA was treated with 30 U of DNase I (QIAGEN) and purified using RNeasy Mini Kit (QIAGEN). A Ribo-Zero Gold rRNA Removal Kit (Illumina) was used to deplete rRNAs from 5 μg of DNase-treated RNAs. The rRNA-depleted RNA was purified by ethanol precipitation and resuspended in 10 μl of nuclease-free water. To remove RNA transcripts carrying a 5' monophosphate moiety (5'-P), 2 μg of rRNA-depleted RNA were treated with 1 U Terminator 5'-Phosphate-Dependent Exonuclease (Epicentre) in the 1x Buffer A in the presence of 40 U RNaseOUT in a 50-μl reaction at 30ºC for 1 h. Samples were extracted with acid phenol-chloroform pH 4.5 (ThermoFisher Scientific), and RNA was recovered by ethanol precipitation and resuspended in 30 μl of nuclease-free water. Next, to remove 5'-terminal phosphates, RNA was treated with 1.5 U CIP (NEB) in 1x NEBuffer 3 in the presence of 40 U RNaseOUT in a 50-μl reaction at 37ºC for 30 min. Samples were extracted with acid phenol-chloroform and RNA was recovered by ethanol precipitation and resuspended in 30 μl of nuclease-free water. To convert 5'-capped RNA transcripts to 5'-monophosphate RNAs ligatable to 5' adaptor, CIP-treated RNAs were mixed with 12.5 U CapClip (Cellscript) and 40 U RNaseOUT in 1x CapClip reaction buffer in a 40-μl reaction, and incubated at 37ºC for 1 h. RNAs were extracted with acid phenol-chloroform, recovered by ethanol precipitation and resuspended in 10 μl of nuclease-free water. To ligate the 5' adapter, the CapClip-treated RNA products were combined with 1 μM 5' adapter oligonucleotide s1086 (5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNN-3'), 1x T4 RNA ligase buffer, 40 U RNaseOUT, 1 mM ATP, 10% PEG 8000 and 10 U T4 RNA ligase 1 in

a 30-µl reaction. The mixtures were incubated at 16°C for 16 h and the reactions were stopped by adding 30 µl of 2x RNA loading dye. The mixtures were separated by electrophoresis on 10% 7 M urea slab gels in 1x TBE buffer and incubated with SYBR Gold nucleic acid gel stain. RNA products migrating above the 5' adapter oligo were recovered from the gel as described[114], purified by ethanol precipitation and resuspended in 10 µl of nuclease-free water.

To generate first strand cDNA, 5'-adaptor-ligated products were mixed with 0.3 µl of 100 µM s1082 oligonucleotide (5'-GCCTTGGCACCCGAGAATTCCANNNNNNNNN-3', N=A/T/G/C) containing a randomized 9-nt sequence at the 3' end, incubated at 65°C for 5 min, and cooled to 4°C. A solution containing 4 µl of 5x First-Strand buffer, 1 µl (40 U) RNaseOUT, 1 µl of 10 mM dNTP mix, 1 µl of 100 mM DTT, 1 µl (200 U) of SuperScript III Reverse Transcriptase and 1.7 µl of nuclease-free water was added to the mixture. Reactions were incubated at 25°C for 5 min, 55°C for 60 min, 70°C for 15 min, and cooled to 25°C. 10 U RNase H was added, the mixtures were incubated 20 min at 37°C and 20 µl of 2x DNA loading solution (PippinPrep Reagent Kit, Sage Science) were added. Nucleic acids were separated by electrophoresis on 2% agarose gel (PippinPrep Reagent Kit, external Marker B) to collect species of ~90 to ~550 nt. cDNA was recovered by ethanol precipitation and resuspended in 20 µl of nuclease-free water. To amplify cDNA, 9 µl of gel-isolated cDNA was added to the mixture containing 1x Phusion HF reaction buffer, 0.2 mM dNTPs, 0.25 µM Illumina RP1 primer (5'-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA-3'), 0.25 µM Illumina index primers RPI3-RPI16 (index primers have the same sequences on 5' and 3' ends, but different on 6-nt sequence that serves as a barcode (underlined); RPI3: 5'-CAAGCAGAAGACGGCATACGAGAT<u>GCCTAA</u>GTGACTGGAGTTCCTTGGCACCCGAG AATTCCA-3'), and 0.02 U/µl Phusion HF polymerase in 30-µl reaction. PCR was performed with

an initial denaturation step of 10 s at 98°C, amplification for 12 cycles (denaturation for 5 s at 98°C, annealing for 15 s at 62°C and extension for 15 s at 72°C), and a final extension for 5 min at 72°C. Amplified cDNAs were isolated by electrophoresis on 2% agarose gel (PippinPrep Reagent Kit, external Marker B) and products of ~180 to ~550 nt were collected. cDNA was recovered by ethanol precipitation and resuspended in 13 μl of nuclease-free water. Barcoded libraries were pooled and sequenced on an Illumina NextSeq platform in high output mode using custom primer s1115 (5'-CTACACGTTCAGAGTTCTACAGTCCGACGATC-3').

### 4.2.2 TSS-seq data analysis:

#### 4.2.2.1 Mapping

Quality control on TSS sequencing library FASTQ files was performed to remove reads with low quality using fastq_quality_filter in the FASTX-Toolkit package (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Cutadapt was then used to remove the 6 nucleotide 5' linker and resulting reads were trimmed from 3' end to 35 nucleotides long[115]. Trimmed reads were mapped to the *S. cerevisiae* R64-1-1 (SacCer3) genome using Bowtie with allowance of no more than two mismatches with suppression of non-uniquely mapped reads, reported in sam files[116]. Uniquely mapped reads were then extracted from sam files using SAMtools and output in bam format[117]. Bam files were then sorted and converted into bed files by SAMtools and BEDTools[117,118]. Customized scripts used bed files to identify the genomic coordinate of the 5' end of each uniquely mapped read. BEDTools was then used to determine pileup (TSS coverage) across the genome, resulting in stranded bedgraph files (https://genome.ucsc.edu/goldenPath/help/bedgraph.html)[118] .

#### 4.2.2.2 TSS correlation

TSS coverage data in bedgraph files were used to examine the correlation between pairwise comparison of TSS libraries. A custom R script was used to filter bedgraph files to examine genome positions with greater than two counts in each library. Log2 transformed TSS counts at the same genomic location in two examined libraries were plotted for all TSS sites to create a heat scatter plot using LSD R package[119], and the Pearson correlation coefficient was calculated. The correlation coefficients deriving from all pairwise comparisons were plotted by a web-based heatmap tool Morpheus (https://software.broadinstitute.org/morpheus/), and clustered by Euclidian distance. Replicates with correlation coefficient greater than 0.85 (0.9 in most recent sequencing data) and the shortest Euclidian distance to each other in the clustering analysis among all the analyzed libraries were recognized as having good sequencing reproducibility and used for downstream analysis.

#### 4.2.2.3 TSS counts table and the heatmap

For each of 5979 selected promoters, TSS usage was examined within 401-nt wide window, spanning 250 nt upstream and 150 nt downstream of the previously annotated median TSS[79]. Using BEDTools and customized R scripts, TSS coverage from the bedgraph files of a library were assigned into the defined windows to generate a 401×5979 TSS count table, with each row representing one of the 5979 promoters, each column represents a promoter position, and the number in each cell representing 5' ends mapping to that position[118]. The count tables generated from two biological replicates were merged into one by aggregating read counts from two. The merged count table was row-wise normalized to get the relative TSS usage in each promoter and visualized by Morpheus. TSS distribution differences were determined by subtracting normalized WT data from normalized mutant data and visualized using heatmaps (Morpheus).

#### 4.2.2.4 Median TSS, the shift and its relationship to the PIC-TSS distance

Data analyses for TSS distributions were based on customized R scripts and results were plotted in Prism 8 (https://www.graphpad.com/scientific-software/prism/) unless otherwise indicated. The median TSS position was defined as the actual TSS containing the 50$^{th}$ percentile of the TSS distribution and was determined for each promoter. "TSS shift" represents the difference in nucleotide of the median TSS position for each promoter between two libraries. The distributions of TSS shifts for all 5979 promoters or selected promoter classes in each library were illustrated by boxplots. Using previously annotated genomic location of the PIC and the median TSS positions in WT yeast cells, the distance between PIC and the median TSS (PIC-TSS distance) was calculated and used to examine the relationship to TSS shifts. PIC-TSS distance was plotted versus TSS shift for all or selected promoters, and linear regression was performed.

#### 4.2.2.5 TSS spread

The spread of TSS, which measures the width of the middle 80% of TSS distribution, was calculated by subtracting positions of 10th percentile and 90th percentile of TSS reads in 401-nt promoter window and add 1. TSS spread of all or selected promoters are shown in boxplot and compared between libraries by performing the ONE WAY ANOVA test. Total reads in each promoter was calculated by summing the reads in 401-nt promoter window and used to measure the expression of the promoter. In each library, the 5979 selected promoters are classified into 1-10 groups by expression levels from low to high. TSS spread was then examined in each expression group and compared between libraries as described above.

#### 4.2.2.6 Differential expression analysis by DESeq2

For each library, total reads in 401-nt promoter window was summed from the count table and the differential expression between libraries was analyzed by DESeq2 (with the alternative hypothesis of $|\log_2 (\text{fold change})| > 1.5$ and a false discovery rate of 0.1)[120]. The $\log_2$ fold changes (log2FC) of the expression between mutant and wild type were taken from the output of DESeq2 analysis and used to examine the dependency on the distance from TSS to PIC. The distance between TSS and PIC (PIC-TSS distance) was measured by subtracting positions of RNA-seq mapped median TSS and our lab's previously ChIP-exo mapped PIC of wild type yeast cells[79]. Filtering was performed to remove promoters having less than 100 of merged reads of wild type libraries and promoters with the calculated PIC-TSS distance in positive values as described in above. Scatter plot was made between $\log_2$FC of the expression and the PIC-TSS distance at all or selected promoters, and the LOcally-WEighted regression Scatterplot Smoother (LOWESS) method was used to fit a curve. Filtered promoters were binned into evenly distributed quintiles depending on the PIC-TSS distance. Binned promoters in each quintile were split into ribosomal and non-ribosomal classes according to annotation in Saccharomyces Genome Database (SGD, https://www.yeastgenome.org/)[121]. For each promoter class, the dependency of $\log_2$FC of expression on PIC-TSS distance was examined in distance quintiles and shown in boxplots.

#### 4.2.2.7 TSS motif analysis

The 401-nt promoter window was the same as used in TSS count table, which spans 250nt upstream and 150nt downstream of the previously annotated yeast wild type median TSS. DNA sequences in 401-nt promoter window were extracted from yeast reference genome *S. cerevisiae* R64-1-1 (SacCer3) and saved as a promoter sequence file. For each TSS that has non-zero mapped reads in the 401nt promoter window, 10-nt sequences from -8 to +1 of the TSS (N-

$_{8}N_{-7}N_{-6}N_{-5}N_{-4}N_{-3}N_{-2}N_{-1}N_{+1}N_{+2}$, TSS at $N_{+1}$) were extracted from the promoter sequence file, as well as the reads number from the TSS count table. Usage of a specific nucleotide (A/T/G/C) at a specific position of 10nt motif (-8 to +1) was calculated by summing the reads of TSS having the same nucleotide type at the same position. For example, "A" usage at -8 position was calculated by summing reads of TSS having an "A" at -8 position. Calculated usage of A/T/G/C at 10nt TSS motif of all or selected promoters was plotted by ggseqlogo[122]. For 64 motifs that are composed of possible nucleotides at -8, -1 and +1 positions of the TSS ($N_{-8}N_{-1}N_{+1}$), the relative usage of each motif at all or selected promoters was calculated as follows. Reads number of TSS that have the same $N_{-8}N_{-1}N_{+1}$ motif was summed up and used to divide the total reads of all TSS to get the percentage of the motif usage. Comparison of the relative usage of 64 motifs shown in percentage was visualized in Morpheus Heatmap. The difference of motif usage between libraries was compared by subtracting the relative usage number and visualized in Morpheus Heatmap.

## 4.3 RESULTS

### 4.3.1 Transcription start site sequencing (TSS-seq) identifies TSS in *S. cerevisiae*

Our TSS-seq method enables identification and quantitative expression analysis of TSS at single-nucleotide resolution. It applies a series of enzymatic treatments to enrich 5' capped Pol II transcripts, followed by cDNA library construction and the sequencing, as illustrated in **Figure 33** (also refer to Materials and Methods 4.2.1).

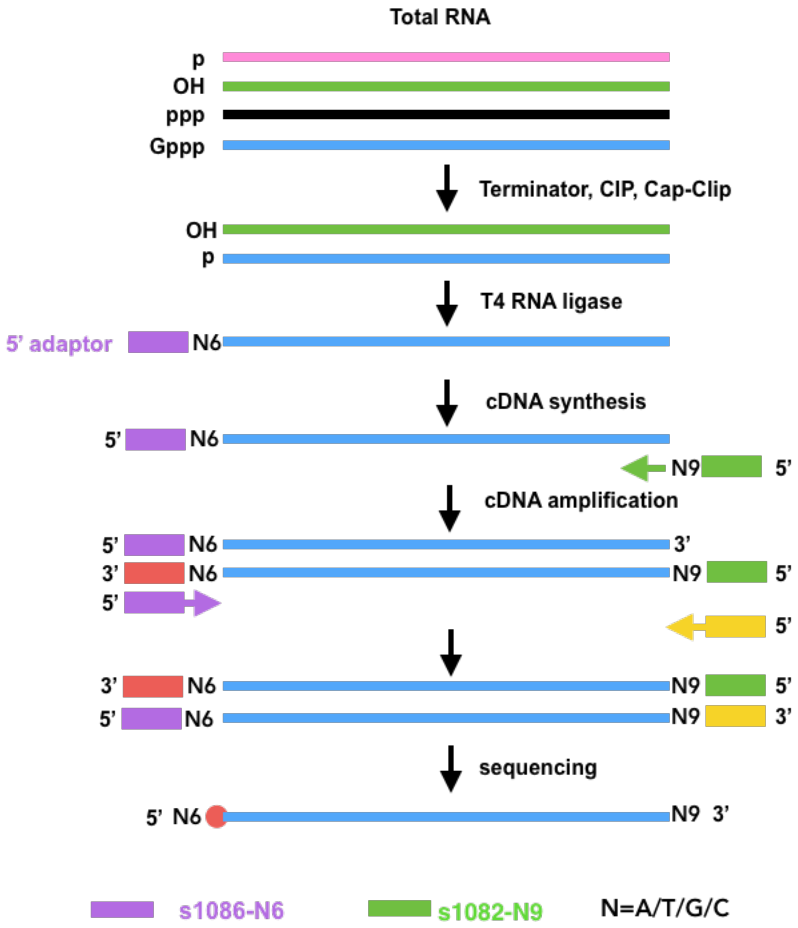



**Figure 33 Flow chart of transcription start site sequencing (TSS-seq)**

Total RNA extracted from yeast cells are treated with terminator, CIP and Cap-Clip sequencially to enrich 5' capped RNA population, followed by 5' adaptor ligation, cDNA conversion, cDNA amplification and the sequencing.

To gain reliable TSS-seq data for TSS usage genome-wide, as applied in other RNA sequencing studies, we performed sequencing on biological replicates and used reproducible data for downstream analysis. In total, we sequenced six *ssl2* mutants, including three upstream shifting alleles L225P, N230I and R636C, and three downstream shifting alleles N230D, D522V and Y750*. Each of them was sequenced in two biological replicates. Due to lower-than-expected correlation between the biological replicates at the single-nucleotide level (though still decent for these types of analyses) during early sequencing batches, we performed a number of rounds of

sequencing. Here, I will discuss the two most recent batches, one batch (batch I) includes Ssl2 WT and the above six *ssl2* mutants, plus Pol II WT, E1103G and H1085Y for comparison purposes. The other batch (batch II) includes Ssl2 WT, N230D and R636C. Data reproducibility analysis on biological replicates in batch I gives us reduced but decent correlation between biological replicates at the single TSS level compared to what our lab had obtained in previous studies (see discussion in below). However, based on the comparison between TSS usage of Pol II mutants in batch I and in our lab's previous research, correlations should be reasonable to give us genome-wide TSS usage information. Data reproducibility analysis on biological replicates in batch II reveals that technical issues giving reduced RNA yields during sequencing library construction has been solved, therefore we are currently repeating all samples for TSS-seq experiments. New sequencing data will be analyzed by a pipeline as discussed in detail below but we do not suspect our present conclusions will be substantially altered.

We found that most of our sequencing libraries have raw total reads of ~20-40 million, we thus proceeded to the next step of data analysis **(Table 10, 11)**. Our TSS-seq used single end sequencing mode and generated RNA reads ~75 bp in length with a 6 bp adapter on 5' end. After removing sequencing reads that don't pass quality control (~3-5% of total reads), the 5' adapter is trimmed so that the first nucleotide of the sequencing reads are the 5' ends of the original RNA transcripts, which are the presumptive TSSs. We then trimmed the reads such that they were 35 nt in length from the 3' end and mapped them to the yeast genome. Mapping statistics show that ~60%-70% processed reads among sequencing libraries were uniquely mapped the yeast genome, the other 30-40% not uniquely mapped reads includes 10% *S. pombe* genome that we added as a potential spike-in control. The mapping statistics and the sequencing reads generated at the intermediate steps are summarized in **Table 10** (batch I) and **Table 11** (batch II).

**Table 10 TSS_seq mapping statistics I**

| Sample ID (Vv) | CKY | Mutant | Number of raw reads | Reads number after filtration | Reads number after adaptor trimming | Reads number after trimming to 35nt | Reads number with at least one reported alignment | Reads number failed to align | Reads number aligned more than once |
|---|---|---|---|---|---|---|---|---|---|
| **1368** | CKY763 | *RPB1*-1 | 38,313,769 | 36224283(-5%) | 36,224,283 | 36,224,250 | 22055000 (60.88%) | 12024165 | 2145085 (5.92%) |
| **1369** | CKY763 | *RPB1*-1 | 21,409,272 | 20645018(-3%) | 20,645,018 | 20,644,984 | 14018721 (67.90%) | 6083853 (29.47%) | 542410 (2.63%) |
| **1370** | CKY764 | *RPB1*-2 | 17,994,836 | 17027023(-5%) | 17,027,023 | 17,026,865 | 9805961 (57.59%) | 6689540 (39.29%) | 531364 (3.12%) |
| **1371** | CKY783 | *rpb1* E1103G-1 | 35,368,590 | 33993357(-3%) | 33,993,357 | 33,993,333 | 20588034 (60.56%) | 10871328 | 2533971 (7.45%) |
| **1372** | CKY784 | *rpb1* E1103G-2 | 42,678,259 | 41178075(-3%) | 41,178,075 | 41,178,071 | 27959117 (67.90%) | 11251676 | 1967278 (4.78%) |
| **1373** | CKY1030 | *rpb1* H1085Y-1 | 28,695,309 | 27490952(-4%) | 27,490,952 | 27,490,926 | 15289432 (55.62%) | 8943832 (32.53%) | 3257662 (11.85%) |
| **1374** | CKY1031 | *rpb1* H1085Y-2 | 46,131,577 | 44540035(-3%) | 44,540,035 | 44,540,033 | 32538690 (73.05%) | 9682498 (21.74%) | 2318845 (5.21%) |
| **1375** | CKY2507 | *SSL2*-1 | 35,186,190 | 34066147(-3%) | 34,066,147 | 34,066,113 | 25054909 (73.55%) | 8115238 (23.82%) | 895966 (2.63%) |
| **1376** | CKY2507 | *SSL2*-1 | 41,178,867 | 39667734(-3%) | 39,667,734 | 39,667,716 | 29174507 (73.55%) | 9433754 (23.78%) | 1059455 (2.67%) |
| **1377** | CKY2508 | *SSL2*-2 | 38,546,168 | 37178226(-3%) | 37,178,226 | 37,178,210 | 24449153 (65.76%) | 11268822 | 1460235 (3.93%) |
| **1378** | CKY2493 | *ssl2* L225P-1 | 37,134,312 | 35778227(-3%) | 35,778,227 | 35,778,123 | 23241722 (64.96%) | 11520369 | 1016032 (2.84%) |
| **1379** | CKY2494 | *ssl2* L225P-2 | 35,306,334 | 33897593(-3%) | 33,897,593 | 33,897,582 | 22748149 (67.11%) | 10154374 | 995059 (2.94%) |
| **1380** | CKY2495 | *ssl2* N230I-1 | 42,344,276 | 40832866(-3%) | 40,832,866 | 40,832,851 | 27661136 (67.74%) | 11761843 | 1409872 (3.45%) |
| **1381** | CKY2496 | *ssl2* N230I-2 | 43,657,254 | 42184187(-3%) | 42,184,187 | 42,184,182 | 30023991 (71.17%) | 9800437 (23.23%) | 2359754 (5.59%) |
| **1382** | CKY2497 | *ssl2* N230D-1 | 35,191,683 | 33179678(-5%) | 33,179,678 | 33,179,674 | 22613318 (68.15%) | 9642226 (29.06%) | 924130 (2.79%) |
| **1383** | CKY2498 | *ssl2* N230D-2 | 28,695,716 | 27322149(-4%) | 27,322,149 | 27,322,146 | 19940126 (72.98%) | 5969918 (21.85%) | 1412102 (5.17%) |
| **1384** | CKY2501 | *ssl2* D522V-1 | 62,950,281 | 59686002(-5%) | 59,686,002 | 59,685,988 | 43023963 (72.08%) | 13402407 | 3259618 (5.46%) |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **1385** | CKY2502 | *ssl2* D522V-2 | 44,030,157 | 41830259(-4%) | 41,830,259 | 41,830,257 | 29534099 (70.60%) | 9808436 (23.45%) | 2487722 (5.95%) |
| **1386** | CKY2503 | *ssl2* R636C-1 | 37,395,884 | 35401057(-5%) | 35,401,057 | 35,401,057 | 25599110 (72.31%) | 7852677 (22.18%) | 1949270 (5.51%) |
| **1387** | CKY2504 | *ssl2* R636C-2 | 45,318,839 | 41107589(-9%) | 41,107,589 | 41,107,575 | 25101453 (61.06%) | 12305199 | 3700923 (9.00%) |
| **1388** | CKY2505 | *ssl2* Y750*-1 | 25,918,261 | 23660295(-8%) | 23,660,295 | 23,660,295 | 13019210 (55.03%) | 8924707 (37.72%) | 1716378 (7.25%) |
| **1389** | CKY2506 | *ssl2* Y750*-2 | 45,773,286 | 42878293(-6%) | 42,878,293 | 42,878,293 | 24058973 (56.11%) | 15512972 | 3306348 (7.71%) |

**Table 11 TSS_seq mapping statistics II**

| Sample ID (Vv) | CKY | Mutant | Number of raw reads | Reads number after filtration | Reads number after adaptor trimming | Reads number after trimming to 35nt | Reads number with at least one reported alignment | Reads number failed to align | Reads number aligned more than once |
|---|---|---|---|---|---|---|---|---|---|
| 1411 | CKY783+ | *rpb1* E1103G-1 | 23,266,618 | 22,459,714 | 22,459,714 | 22,459,706 | 15880638 (70.71%) | 4475854 (19.93%) | 2103214 (9.36%) |
| 1412 | CKY784+ | *rpb1* E1103G-2 | 54,844,812 | 53,189,810 | 53,189,810 | 53,189,786 | 37479245 (70.46%) | 10961335 (20.61%) | 4749206 (8.93%) |
| 1413 | CKY1030 | *rpb1* H1085Y-1 | 48,715,215 | 47,053,481 | 47,053,481 | 47,053,450 | 33940704 (72.13%) | 9339982 (19.85%) | 3772764 (8.02%) |
| 1414 | CKY1031 | *rpb1* H1085Y-2 | 14,241,836 | 13,792,827 | 13,792,827 | 13,792,823 | 10081791 (73.09%) | 2419153 (17.54%) | 1291879 (9.37%) |
| 1415 | CKY2497 | *ssl2* N230D-1 | 38,434,842 | 37,265,975 | 37,265,975 | 37,265,973 | 25776540 (69.17%) | 7634536 (20.49%) | 3854897 (10.34%) |
| 1416 | CKY2498 | *ssl2* N230D-2 | 31,508,730 | 30,557,966 | 30,557,966 | 30,557,965 | 21542466 (70.50%) | 6333628 (20.73%) | 2681871 (8.78%) |
| 1417 | CKY2503 | *ssl2* R636C-1 | 38,471,262 | 37,323,306 | 37,323,306 | 37,323,305 | 26621619 (71.33%) | 7305653 (19.57%) | 3396033 (9.10%) |
| 1418 | CKY2504 | *ssl2* R636C-2 | 27,370,937 | 26,542,375 | 26,542,375 | 26,542,370 | 18426612 (69.42%) | 5668406 (21.36%) | 2447352 (9.22%) |
| 1419 | CKY2507 | *SSL2*-1 | 24,229,338 | 23,395,647 | 23,395,647 | 23,395,644 | 16307291 (69.70%) | 5032466 (21.51%) | 2055887 (8.79%) |
| 1420 | CKY2508 | *SSL2*-2 | 36,981,869 | 35,792,106 | 35,792,106 | 35,792,106 | 24854983 (69.44%) | 8202783 (22.92%) | 2734340 (7.64%) |
| 1421 | CKY2508 | *SSL2*-2 | 26,512,465 | 25,673,647 | 25,673,647 | 25,673,633 | 18145065 (70.68%) | 5781407 (22.52%) | 1747161 (6.81%) |
| 1422 | CKY2508 | *SSL2*-2 | 32,497,910 | 31,443,235 | 31,443,235 | 31,443,233 | 22102637 (70.29%) | 7031526 (22.36%) | 2309070 (7.34%) |

The positions and counts of the 5' ends of the uniquely mapped reads were then extracted using our lab-developed scripts and used to assess the reproducibility between biological replicates. The Pearson correlation coefficient between replicates was calculated based on the reads counts at single TSS positions. We found that most replicates in TSS-seq batch I have the Pearson correlation coefficient ranges from ~0.6 to ~0.9 **(Figure 34)**. The TSS correlation heatmap shows that replicates with a Pearson correlation coefficient number ~0.64 (e.g. *rpb1* H1085Y) is more scattered than replicates with a higher coefficient number (e.g. *ssl2* D522V), indicating a poorer correlation between replicates with lower coefficient number.

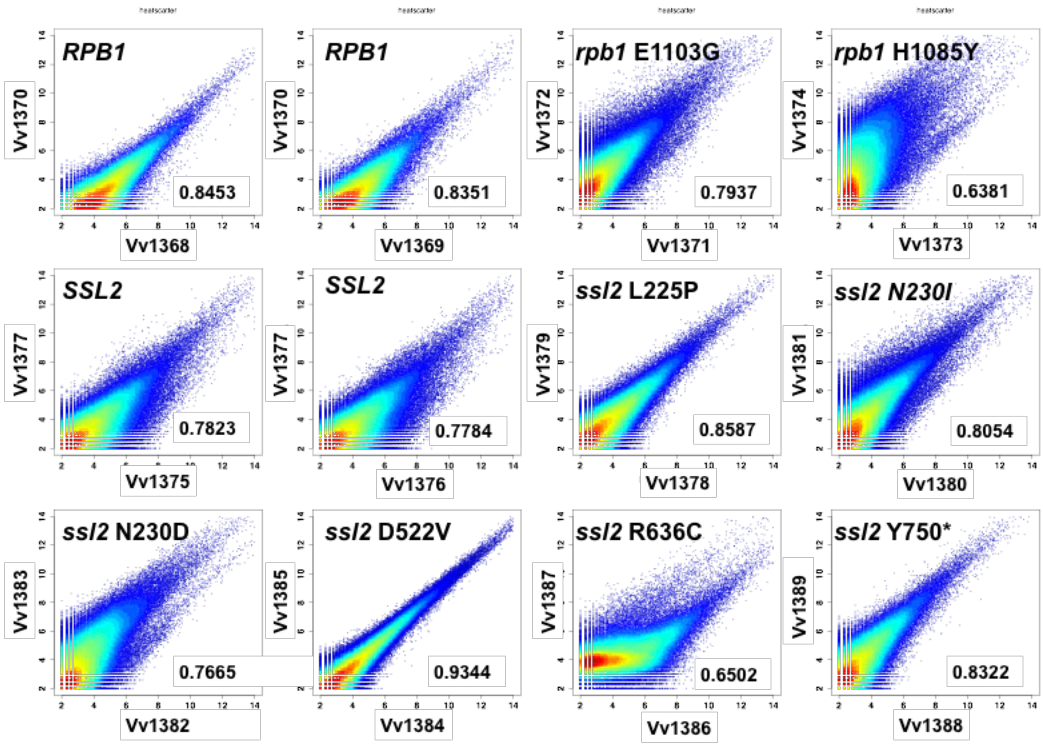



**Figure 34 Correlation of reads number between replicates at single TSS level (batch I)**

Scatter plot showing the correlation of the reads number at individual TSS between replicates in TSS-seq batch I. Replicates with higher Pearson correlation coefficient number show more focued correlation (e.g. *ssl2* D522V), whereas replicates with lower Pearson correlation coefficient number show more scattered correlation (e.g. *rpb1* H1085Y). In scatter plot, each dot represents a TSS, and the x-axis represents TSSs' reads number in biological

replicate 1 and y-axis represents the correspondent TSS's reads number in biological replicate 2. Individual TSS is showed in blue and the overlayed TSSs are shown by griendent color from cyan, green, yellow to red to indicate increased TSS density.

In TSS-seq batch II, all replicates show higher than 0.92 Pearson correlation coefficient and greatly correlated TSSs reads numbers **(Figure 35)**. We set this level of correlation as a standard to qualify reproducible biological replicates. For each wild-type or mutant strain, if the biological replicates of TSS-seq have a calculated Pearson correlation coefficient higher than 0.9 and a consistent visual correlation between the reads number of individual TSS in the heatmap, we proceed to downstream analysis.

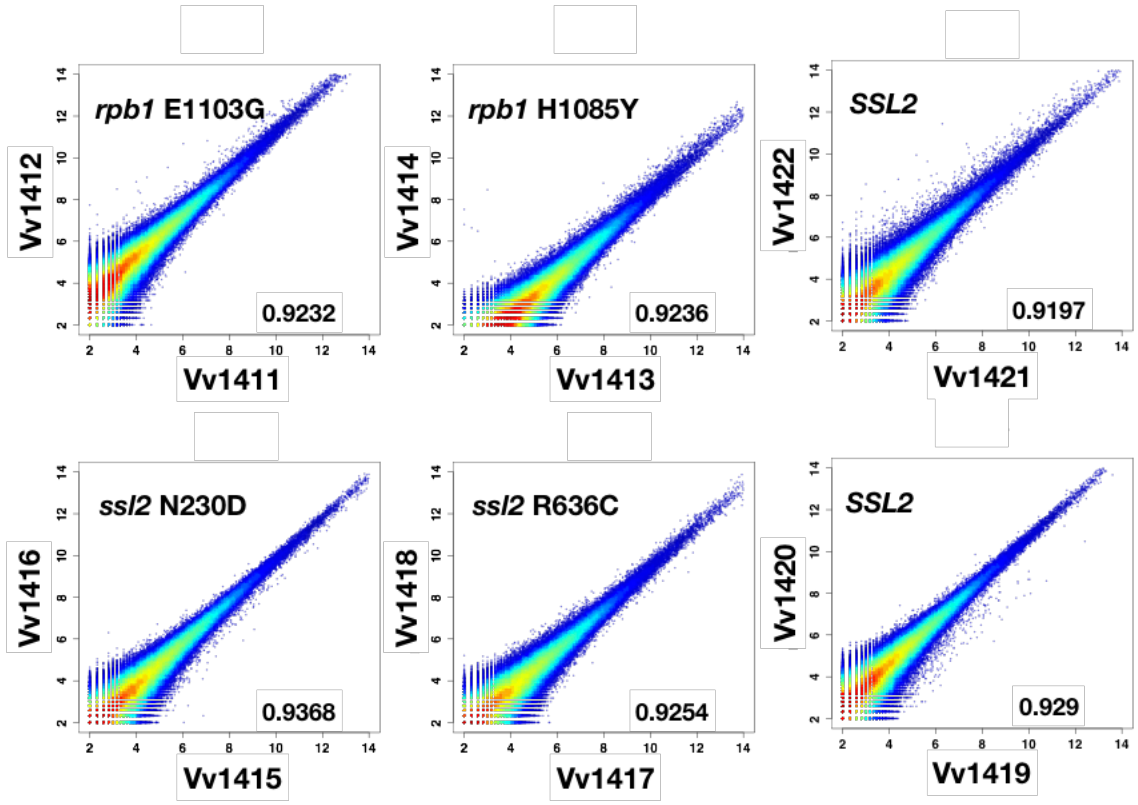



**Figure 35 Correlation of reads number between replicates at single TSS level (batch II)**

Scatter plot showing the correlation of the reads number at individual TSS between replicates in TSS-seq batch II. All the replicates in batch II showed strong correlation as indicated by the high correlation coefficient number and the focused correlation plot.

Our lab's previous study found that clustering of correlation coefficients among libraries could distinguish Pol II mutants into GOF and LOF groups[79]. Here, we ask if *ssl2* mutant classes could also be distinguished by the TSS based correlation coefficients. To answer this, we performed hierarchical clustering on Pearson correlation coefficients among the TSS-seq libraries that include both *ssl2* and *rpb1* samples. We noticed that mutant classes are again distinguished based on the 5' reads number at TSS level **(Figure 36)**. For example, Ssl2 WT libraries are clustered in group Ia, downstream shifting *ssl2* R636C libraries are clustered into group Ib, upstream shifting libraries *ssl2* N230D in group IIa, *rpb1* upstream shifting and downstream libraries in group IIb and III respectively. Interestingly, two alleles from different genes that both shift TSSs upstream, *ssl2* N230D and *rpb1* E1103G, and are necessarily derived from different strains, have a closer Euclidean distance than the distance between two alleles from the same strain background but shift TSS distribution to opposite directions, *rpb1* E1103G and *rpb1* H1085Y. In hierarchical clustering of the Pearson correlation coefficients, *ssl2* N230D and *rpb1* E1103G are classified into group II, whereas *rpb1* H1085Y in group III, *ssl2* R636C and WT in group I. Suggesting that there is some degree of similarity between upstream shifting alleles for changing TSS distributions.

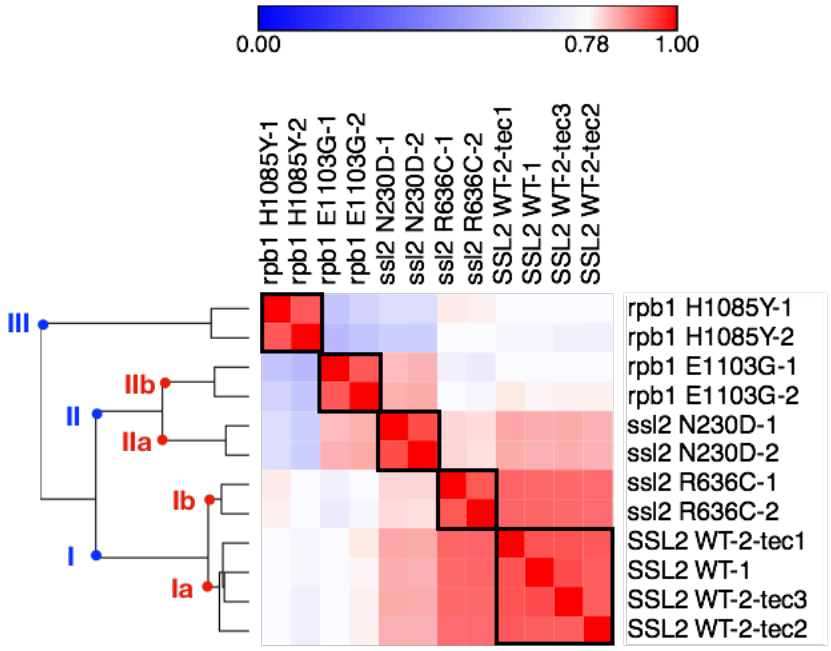



**Figure 36 Matrix heatmap and the hierarchical clustering of Pearson correlation coefficients between TSS-seq libraries**

Matrix showing the Pearson correlation coefficients among TSS-seq libraties. Ssl2 and Rpb1 TSS-seq ibraries are clustered into group I, II and III or I, Ia, Ib, IIa, IIb and III based on the Euclidean distance of Pearson correlation coefficients between libraries. Clustering distinguishes different mutant strains into Ssl2 WT, *ssl2* mutant and *rpb1* mutants. Upstream shifting alleles *ssl2* N230D and *rpb1* E1103G from different strain background show a closer distance in clustering than distance between alleles of *rpb1* H1085Y and *rpb1* E1103G that shift TSS distribution to opposite directions.

We additionally examined mapped sequencing reads at *ADH1* to assess if the TSS usage detected by TSS-seq at individual promoter is consistent with the corresponding TSS usage detected by the primer extension. We observed that the TSS distribution at *ADH1* promoter detected by TSS-seq shows very similar pattern to PE detected TSS usage **(Figure 37)**. In the Ssl2 WT strain, PE data shows that upstream major TSS is slightly more used than the downstream one, quantitatively showing as 63.9M counts and 43.2M counts by phosphorimaging TSS signal in the

three replicates combined sample, and ~1.48-fold TSS usage ratio between the two major sites **(Table 12)**. TSS-seq detected TSS usage at *ADH1* similarly exhibits more upstream usage than the downstream one, showing as 137470 and 91346 mapped reads from the two biological replicates merged data at the two major sites and ~1.50-fold ratio between the two **(Table 13)**. In addition, comparison between TSS-seq detected and PE detected TSS usage at *ADH1* promoter in *ssl2* N230D and R636C also reveals very similar TSS usage patterns **(Table 12, 13)**.

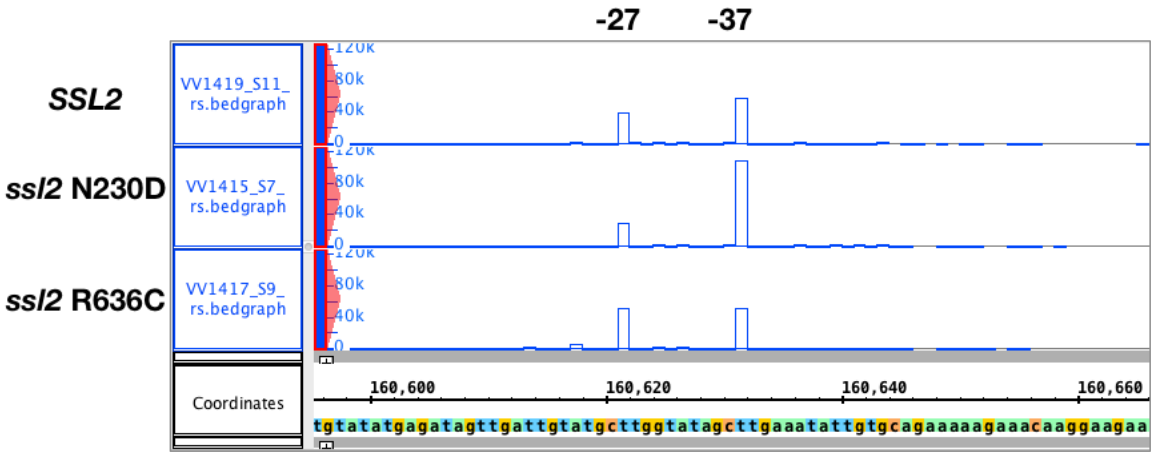



**Figure 37 TSS-seq detected TSS usage is similar to PE detected TSS usage at *ADH1***

Shown is the TSS-seq detected TSS usage at the *ADH1* promoter. TSS-seq mapped sequencing reads were saved in bedgraph file, loaded on IGB and shown in above bargraph[123]. In *SSL2* WT library, upstream TSS (-37) is slightly more used than the downstream one (-27). In N230D, TSS usage at downstream site is reduced, in contrast, TSS usage at downstream site in R636C is increased. These TSS usage patterns of the correspondent WT and mutants are similar to what we have observed in PE assay **(Figure 30)**.

**Table 12 Quantification of TSS usage at *ADH1* promoter detected by primer extension**

| PE | TSS | Replicate 1 | Replicate 2 | Replicate 3 | Sum | -37/-27 ratio |
|---|---|---|---|---|---|---|
| **WT** | -37 | 24,306,582 | 15,720,209 | 23,837,852 | 63,864,643 | 1.47986974 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | -27 | 17,161,455 | 10,624,424 | 15,369,705 | 43,155,584 | |
| **N230D** | -37 | 25,740,733 | 28,450,201 | 29,983,625 | 84,174,559 | 4.59995153 |
| | -27 | 5,689,265 | 6,256,231 | 6,353,514 | 18,299,010 | |
| **R636C** | -37 | 23,683,614 | 24,308,134 | 17,444,367 | 65,436,115 | 0.96710898 |
| | -27 | 24,397,036 | 25,482,629 | 17,781,908 | 67,661,573 | |

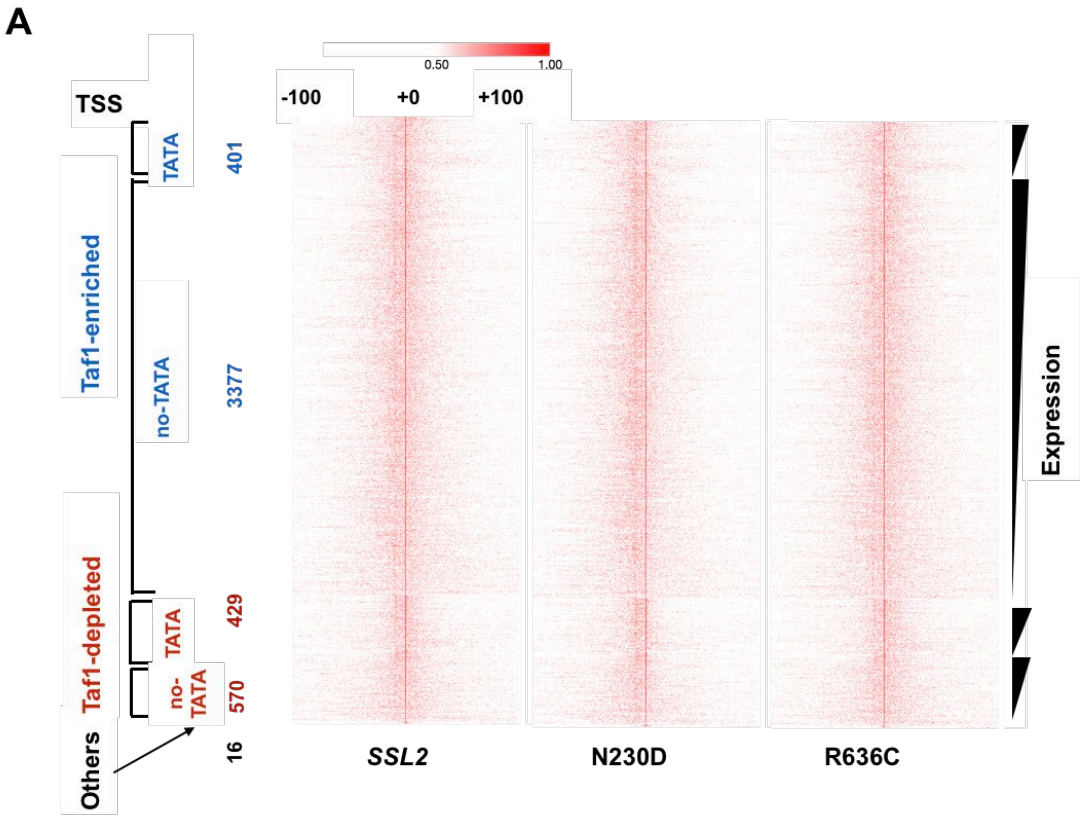
**Table 13 TSS-seq mapped reads number at *ADH1* promoter**

| TSS-seq | TSS | Replicate 1 | Replicate 2 | Sum | -37/-27 ratio |
|---|---|---|---|---|---|
| **WT** | -37 | 56,759 | 80,711 | 137,470 | 1.50493727 |
| | -27 | 38,832 | 52,514 | 91,346 | |
| **N230D** | -37 | 107,709 | 87,066 | 194,775 | 3.75687144 |
| | -27 | 28,661 | 23,184 | 51,845 | |
| **R636C** | -37 | 50,610 | 39,950 | 90,560 | 1.02415634 |
| | -27 | 50,528 | 37,896 | 88,424 | |

### 4.3.2 *ssl2* alleles shift promoter TSS distributions

To examine TSS promoter distributions genome-wide in WT and *ssl2* mutant strains, mapped 5' end read numbers at each nucleotide position of a 401 bp window, which spans 250 bp upstream and 150 bp downstream of previously annotated wild type median TSSs, were window-normalized for 5979 yeast promoters and displayed in heatmaps **(Figure 38A)**. Visualization of TSS distributions in WT yeast shows that our TSS-seq mapped 5' reads are distributed around

previously annotated WT median TSSs, indicating the consistency between TSS-seq experiments. Comparison of TSS distributions between WT and *ssl2* N230D, a hypothetical processivity LOF allele, reveals distinct decrease of TSS reads downstream of the annotated WT median TSS across essentially all promoters, including TATA-containing or TATA-less, Taf1-enriched or depleted promoters. In contrast, TSS distributions of *ssl2* R636C, a processivity GOF allele, show relative increase in TSS reads at downstream of median TSS positions, which is consistent with our previous observation for this allele increasing TSS usage downstream at *IMD2* and *ADH1* promoters. Next, we measured the changes of TSS usage at every TSS between the mutant and wild-type strain by subtracting promoter-normalized reads number between the two and showed in a median TSS aligned window **(Figure 38B)**. We observed decreased TSS signal, showing in cyan, downstream of the originally annotated WT median TSS in *ssl2* N230D mutant. In contrast, substantially increased TSS usage, presenting in orange, is observed at downstream sites in *ssl2* R636C mutant.


A
0.50
1.00
-100
+0
+100
TSS
TATA
Taf1-enriched
no-TATA
Taf1-depleted
TATA
no-TATA
Others
401
3377
429
570
16
Expression
SSL2
N230D
R636C

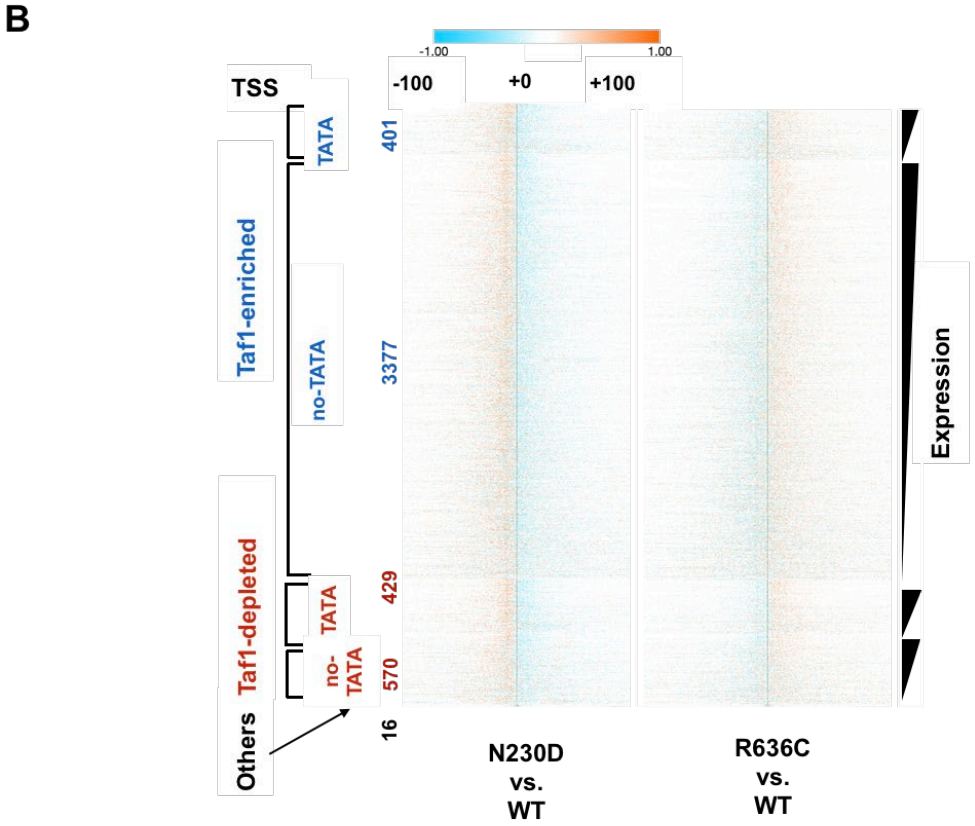

B
-1.00
1.00
-100
+0
+100
TSS
TATA
Taf1-enriched
no-TATA
Taf1-depleted
TATA
no-TATA
Others
401
3377
429
570
16
Expression
N230D
vs.
WT
R636C
vs.
WT

**Figure 38 *ssl2* mutants shift TSS ditribution genome-wide**

**(A)** Heatmap showing TSS reads distributions at WT median TSS centered windows in Ssl2 WT, N230D, and R636C. Each row represents a promoter, and there are 4793 promoters that pass the ≥100 reads in WT filter. Promoter classes are labeled on the left and are ordered according to expression level in WT. Each column represents a TSS site within 201 bp median TSS centered window. Reads at each TSS site are normalized to the total reads for the corresponding promoter (401 bp window) and are shown (201 bp windows are showed in this figure) by a color scheme ranging from white to red. **(B)** Heatmap of TSS reads differences between WT and mutants within median TSS-centered windows. The promoter-normalized reads difference between mutant and WT are shown by a color scheme ranging from cyan-white-orange, with the cyan representing relative decrease in reads compared to WT and orange representing relative increase in reads compared to WT. Compared to WT, N230D decreases relative TSS reads downstream. In contrast, R636C increases relative TSS reads downstream compared to WT.

### 4.3.3 The magnitude of TSS shift in *ssl2* alleles is smaller than in Pol II activity mutants

To quantify the changes of TSS usage, we measured the TSS shift between WT and mutant strains by subtracting the median TSS positions of the 401 bp promoter window between the two[79]. We measured median TSS shifts across promoters divided into classes and subclasses based on presence/absence of TATA element or enrichment/depletion of Taf1[124]. As expected, the directions of median TSS shifts for all promoters and promoter subclasses were consistent with TSS shifts observed at *ADH1* promoters: a negative value generated by subtracting WT median TSS position from the mutant position indicating an upstream TSS shift and was observed in all tested *ssl2* processivity LOF alleles, and a positive value indicating a downstream TSS shift was observed in *ssl2* processivity GOF alleles **(Figure 39A-39F)**. We observed that the magnitude of TSS shift is consistent with the strengths of putative TSS-shift dependent growth phenotypes. For example, *ssl2* downstream shifting alleles N230I, L225P and R636C, from the left to right, show

weaker $His^+$ phenotypes in genetic tests, while also show a gradient in TSS shift magnitudes across promoters. Similarly, *ssl2* upstream shifting alleles N230D, D522V and Y750* from the left to the right show weaker $MPA^S$ phenotypes in genetic tests, their correspondent genome-wide TSS shift magnitudes decrease in similar fashion. Notably, the extents of TSS shifts in *ssl2* alleles are less than for Pol II activity mutants, indicating a more dramatic effect of Pol II's catalytic activity on TSS distribution.

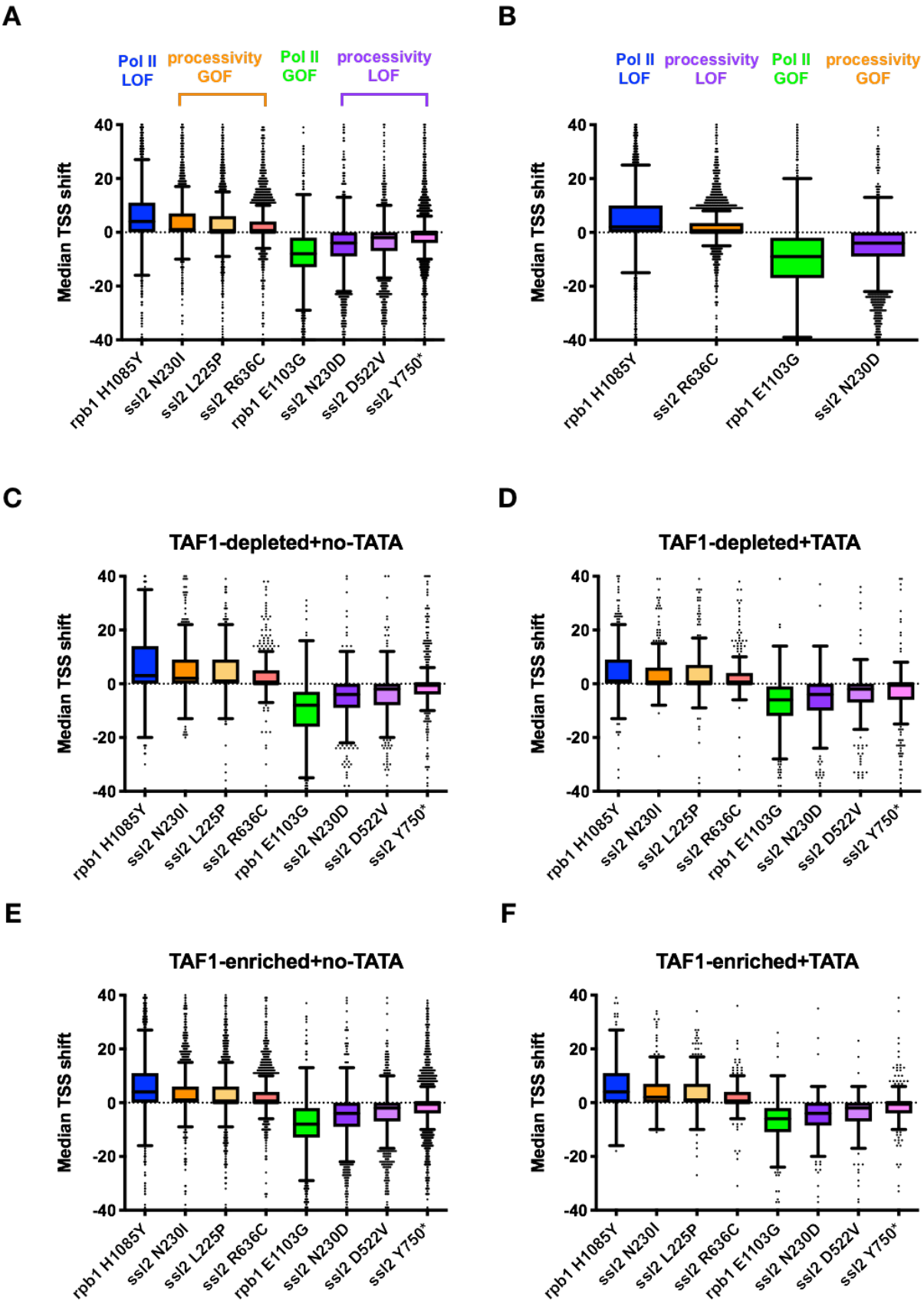

A
Pol II LOF
processivity GOF
Pol II GOF
processivity LOF
Median TSS shift
40
20
0
-20
-40
rpb1 H1085Y
ssl2 N230I
ssl2 L225P
ssl2 R636C
rpb1 E1103G
ssl2 N230D
ssl2 D522V
ssl2 Y750*
B
Pol II LOF
processivity LOF
Pol II GOF
processivity GOF
Median TSS shift
rpb1 H1085Y
ssl2 R636C
rpb1 E1103G
ssl2 N230D
C
TAF1-depleted+no-TATA
Median TSS shift
D
TAF1-depleted+TATA
Median TSS shift
E
TAF1-enriched+no-TATA
Median TSS shift
F
TAF1-enriched+TATA
Median TSS shift

**Figure 39 TSS shifts in *ssl2* mutants are not as strong as TSS shifts in Pol II mutants**

**(A-F)** The median TSS shift of *ssl2* and Pol II mutants in all promoter class, TATA-containg/non-TATA or Taf1-enriched/depleted promoter classes. The degree of TSS shift is correlated with the strength of genetic tested growth phenotypes. Stronger $His^+$ allele show greater TSS shift downstream. Alleles of N230I, L225P and R636C, from left to right showing weaker His+ phenotype and lesser degree of TSS shift. Stronger $MPA^S$ allele show greater TSS shift upstream. Alleles of N230D, D522V and Y750*, from left to right showing weaker $MPA^S$ phenotype and lesser degree of TSS upstream shift.

### 4.3.4 Effects on the width of TSS distribution are distinct between Pol II and *ssl2* alleles

To evaluate the effect of *ssl2* alleles on the distance of scanning, the width between positions of the $10^{th}$ and $90^{th}$ percentiles of the TSS signal at each 401 bp promoter window was determined, which measures the spread of 80% of TSS. We observed that Pol II mutants of both mutant types, catalytic GOF or LOF, show increased TSS spread independent of promoter class, suggesting Pol II alleles' activity in activating not commonly used TSS up or downstream sites are able to make TSS distributions wider than WT **(Figure 40)**. Notably, the width of TSS spread in Pol II GOF is less than in Pol II LOF mutants. A possible explanation for this observation is that there is more constraint at upstream site of initiation, for example, an assembled pre-initiation complex. Alternatively, activation of upstream sites through increased efficiency across all sites would result in Pol II flux running out faster, so that Pol II GOF expanded TSS distributions are narrower than Pol II LOF expanded TSS distributions. As we mentioned earlier, *ssl2* processivity LOF alleles reduce downstream TSS and are predicted to have narrower scanning windows. Indeed, a smaller TSS spread is observed in *ssl2* LOF alleles compared to WT (except for Y750*, which behaves differently from other LOF alleles in a number of aspects, which needs a further discussion). Moreover, consistent with our hypothesis that *ssl2* processivity GOF alleles drive Pol

II scanning further and make scanning windows wider, we observed wider TSS spread in *ssl2* processivity GOF alleles compared to WT. We observed that the magnitude of increased TSS spread between Pol II alleles and *ssl2* downstream shifting alleles showed no clear difference. For example, the strongest Pol II LOF mutants of *rpb1 H1085Y*, which shifts TSSs downstream, had almost the same TSS spread as the potential *ssl2* processivity GOF alleles that also shifts TSSs downstream. We hypothesize that one possible limit for Pol II LOF allele is Ssl2's processivity.

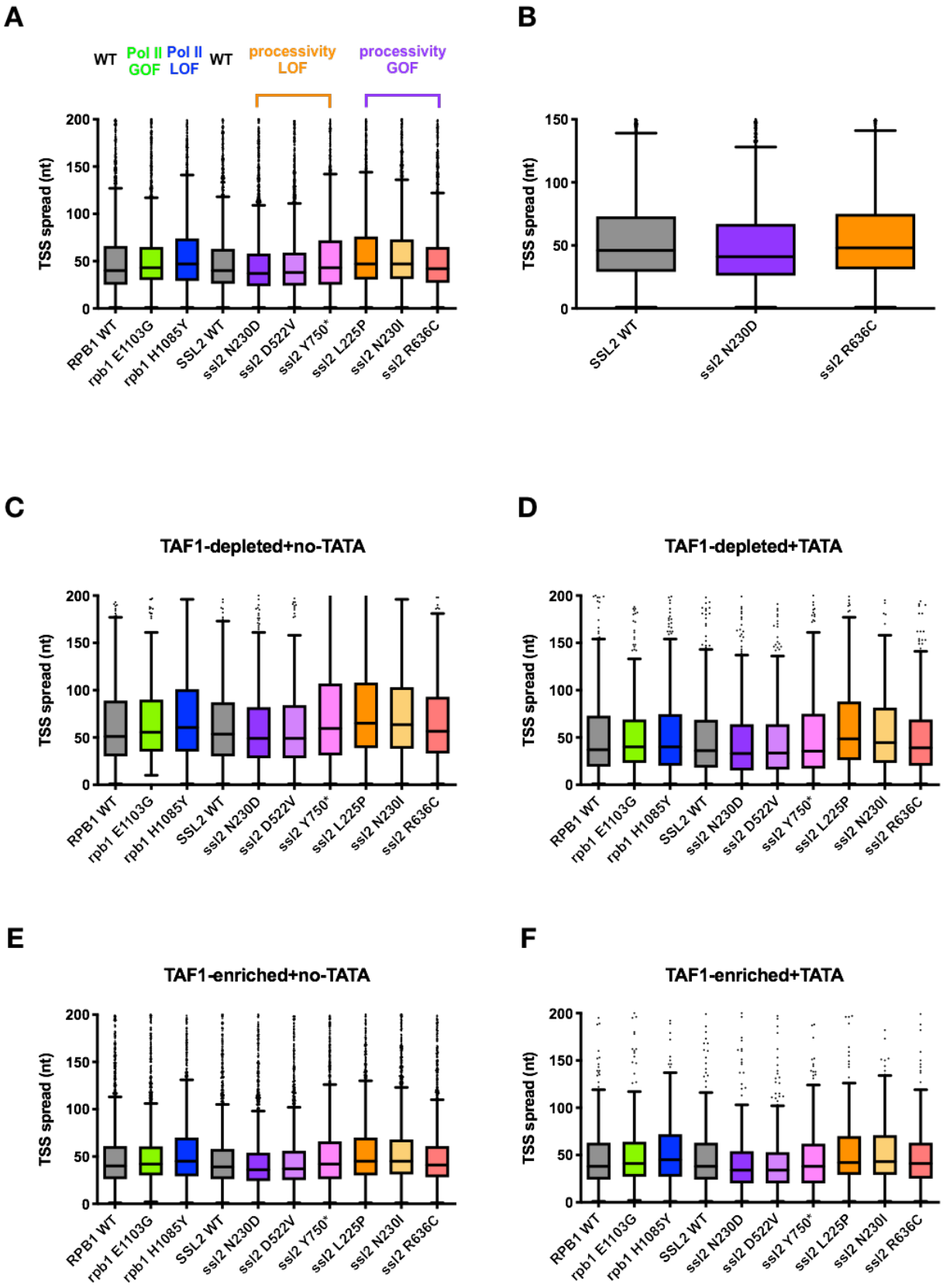

A
WT
Pol II GOF
Pol II LOF
WT
processivity LOF
processivity GOF
TSS spread (nt)
RPB1 WT
rpb1 E1103G
rpb1 H1085Y
SSL2 WT
ssl2 N230D
ssl2 D522V
ssl2 Y750*
ssl2 L225P
ssl2 N230I
ssl2 R636C
B
TSS spread (nt)
SSL2 WT
ssl2 N230D
ssl2 R636C
C
TAF1-depleted+no-TATA
TSS spread (nt)
D
TAF1-depleted+TATA
TSS spread (nt)
E
TAF1-enriched+no-TATA
TSS spread (nt)
F
TAF1-enriched+TATA
TSS spread (nt)

**Figure 40 Hypothetical *ssl2* processivity alleles affect the TSS distribution width**

Pol II activity GOF and LOF alleles increase spread of TSS distributions through increasing TSS usage at upstream or downstram sites. Hypothetical *ssl2* processivity GOF alleles also show increased spread of TSS distribution relative to WT. In constrast, hypothetical *ssl2* processivity LOF alleles exhibit decreased spread of TSS distribution.

### 4.3.5 Distinct relationships between the degree of TSS shift and the PIC-TSS distance in Pol II and *ssl2* mutants

In our prior study, we examined the distance between the PIC and the median TSS by measuring the space between ChIP-exo mapped PIC components Ssl2/Sua7, and 5'-RNA sequencing mapped median TSS of WT yeast strain[79]. We found that most yeast promoters have a putative PIC-TSS distance between 30-120nt. We now investigated the relationship between PIC-TSS distance and the degree of TSS shift in *ssl2* mutants. This analysis asks if putative scanning distance has a relationship with promoter sensitivity to *ssl2* mutants. As a control, we first examined the relationship of PIC-TSS distance and the extent of TSS shift in Pol II catalytic activity mutants. Consistent with our expectations, the longer the PIC-TSS distance is, the greater TSS shift is observed in Pol II GOF allele **(Figure 41)**. This fits with our assumption that when TSS is further away from PIC, there is more room for Pol II GOF allele to activate and shift upstream. Conversely, Pol II LOF allele shows weaker shift with longer PIC-TSS distance by having less room to shift further downstream when processivity is predicted to limit the scanning window. In addition, the correlation between TSS shift and PIC-TSS distance, which is measured by the slope here, is the same in Pol II GOF and LOF allele. This is in agreement with the degree to which their catalytic activities deviate from WT though in opposing directions. We observed that the extent of TSS shift in *ssl2* processivity LOF allele (N230D) is positively correlated to the

PIC-TSS distance, in a way similar to Pol II GOF allele. The further TSS is away from the PIC, the greater shift is observed in *ssl2* processivity LOF allele, indicating that TSS further downstream of the PIC is more sensitive to the reduced processivity. Moreover, the degree of TSS shift in *ssl2* processivity GOF allele (R636C) is not sensitive to the distance of PIC-TSS, suggesting the magnitude of TSS shift is not affected by how far TSS is away from the PIC. We hypothesize that there might be some coupling between Pol II and Ssl2 for initiation. A simple model is that processivity is downstream of initiation activity, in the beginning Pol II and some other factors determine where to initiate and processivity determines how far to scan afterwards.

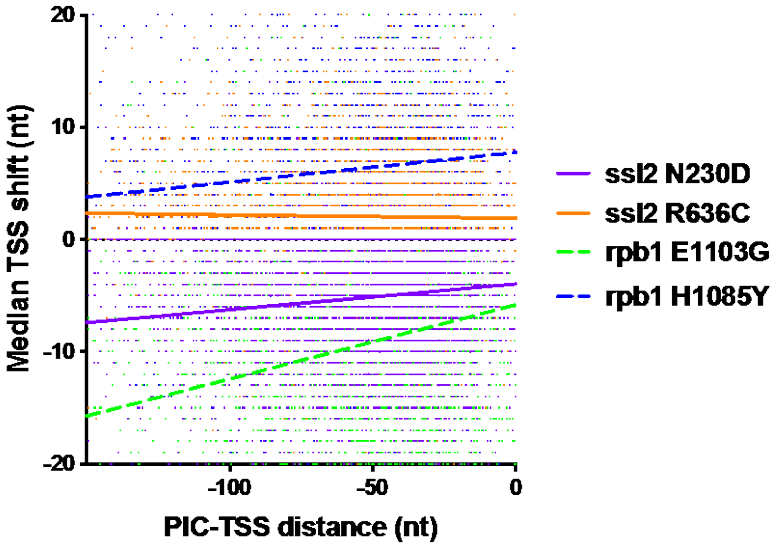



**Figure 41 TSS further downstream of the promoter window is sensitive to the reduced processivity**

*ssl2* N230D, the hypothetical processivity LOF allele shows greater shift at longer PIC-TSS distance. The degree of TSS shift in *ssl2* R636C, the processivity GOF allele, is not sensitive to the PIC-TSS distance.

### 4.3.6 Distinct relationships between differential promoter expression and PIC-TSS distance in *ssl2* upstream shifting and downstream shifting mutants

We also examined the relationship between PIC-TSS distance and differential promoter expression in *ssl2* mutants to ask if *ssl2* mutants were sensitive to PIC-TSS distance for differential promoter expression. We observed that differential promoter expression was not sensitive to the PIC-TSS distance in putative processivity GOF allele *ssl2* R636C, whereas it was negatively correlated with increased PIC-TSS distance – very slightly for Taf1-enriched promoters and more for Taf1-depleted promoters – in the hypothetical processivity LOF *ssl2* N230D **(Figure 42)**. This result suggests a potential for reduced scanning processivity to result in reduced overall initiation efficiency, which would be predicted to affect promoters with distal TSSs more than promoters with shorter PIC-TSS distances.

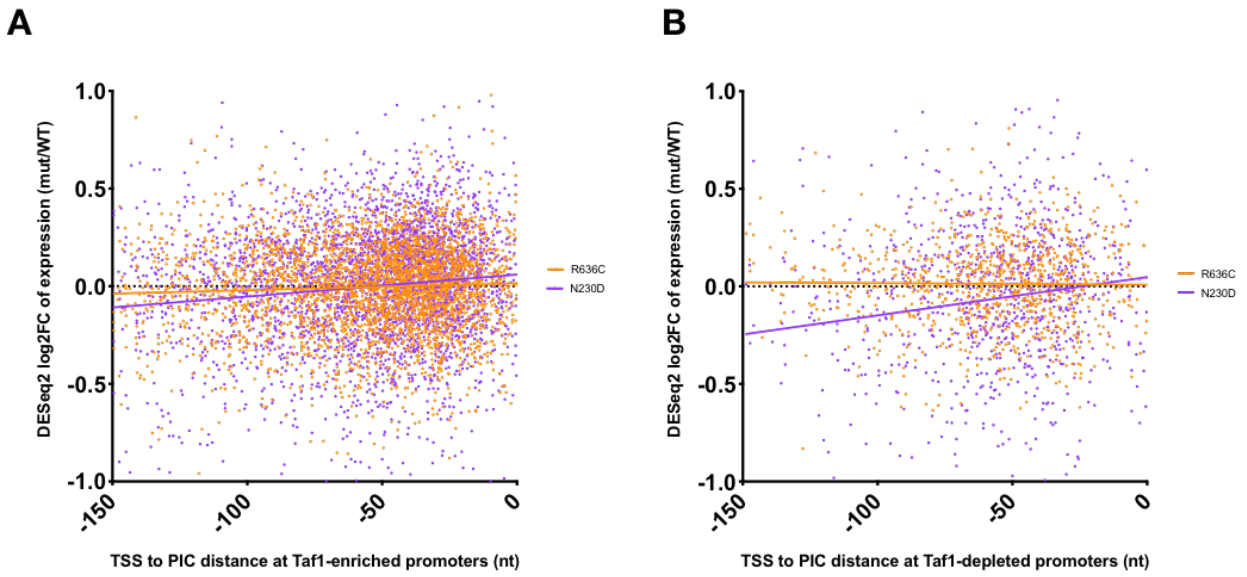



**Figure 42 The relationship between the differential promoter expression and the PIC-TSS in *ssl2* mutants**

Scatterplot and linear regression showing that the differential promoter expression in *ssl2* upstream shifting allele N230D is negatively correlated to the PIC-TSS distance in both Taf1-enriched and Taf1-depleted promoters. Linear

regression in *ssl2* N230D allele shows that Taf1-depleted promoters have stronger negative PIC-TSS distance to differential expression correlation than the Taf1-enriched promoters. Linear regression in *ssl2* downstream shifting allele R636C shows that the differential promoter expression is not correlated to the PIC-TSS distance for either Taf1-enriched or Taf1-depleted promoters.

### 4.3.7 Sequence preference in *ssl2* mutants

Our lab's previous study and others found that Pol II has a preference for using YR (pyrimide-purine) motif at -1 and +1 positions of the non-template strand, as well as an A nucleotide at the -8 position. These preferences are also observed in our TSS-seq data, in both WT and *ssl2* mutants **(Figure 43)**. Usage of a specific nucleotide (A/T/G/C) at a specific position of the 10nt motif (-8 to +1) surrounding a TSS was calculated as the percentage of the aggregated promoter TSS reads at that position and plotted by ggseqlogo[122] (see Materials and Methods). We noticed that the preference for using "A" at -8 position is relatively reduced in *ssl2* upstream shifting allele N230D, whereas there is no obvious difference between WT and downstream shifting allele R636C. We also examined motif-based TSS usage by separating TSS motifs into 64 groups according to the sequence specificity at -8, -1 and +1 positions **(Figure 43B)**. Our lab's previous study found that the top four motifs are AYR, and the next preferred motifs are within BYR[79]. This trend was also found in our TSS-seq data analysis across all the sequenced yeast strains, including WT and the *ssl2* mutants. The change of the -8-1+1 motif usage in *ssl2* mutant was examined by comparing the promoter TSS usage between WT and the mutant. We found that two of the upstream shifting *ssl2* alleles showed decreased preference for the two most preferred TSS motifs but increased preference at BYR motifs, similar to our prior observation in Pol II alleles **(Figure 43C)**. In addition, three downstream shifting *ssl2* alleles L225P, N230I, R636C and

another upstream shifting allele Y750* exhibited increased preference on the most preferred TSS motif ACA.

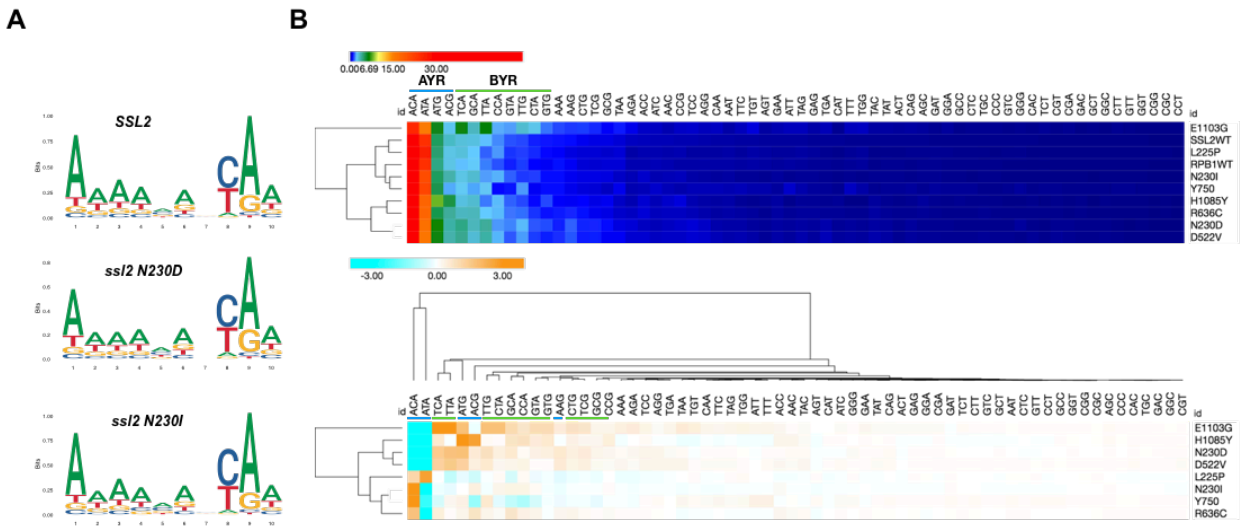



**Figure 43 TSS motif usage at -8, -1 and +1 positions in *ssl2* mutants**

**(A)** WT and *ssl2* mutants show preference for uing AYR motifs. Compared to WT, the preference for using an A at -8 is decreased in *ssl2* upstream shiting N230D allele, wherease there is no obvious difference between WT and *ssl2* downstream shifting N230I mutant. **(B)** TSS motif preference is altered in *ssl2* and Pol II mutants. Top heatmap showing that the most preferred TSS motif is AYR and the next prevalent motifs are within BYR. Bottom panel shows that the preference for using $A_{-8}C_{-1}A_{+1}$ and $A_{-8}T_{-1}A_{+1}$ is decreased in Pol II mutants E1103G, H1085Y, and in upstream shifting *ssl2* alleles N230D and D522V. In contrast, the preference for using $A_{-8}C_{-1}A_{+1}$ is decreased in *ssl2* alleles L225P, N230I, R636C and Y750*.

## 4.4 DISCUSSION

Using genome-wide transcription start site sequencing, we have found that *ssl2* mutants alter TSS usage at all promoter classes. Mutants of *ssl2* N230D, D522V and Y750* that shift TSS

distribution at the model gene *ADH1* also shift TSS distribution upstream genome-wide. Consistent with our hypothesis that these alleles are defective in TFIIH processivity for controlling Pol II scanning downstream. In contrast, alleles of L225P, N230I and R636C shift TSS distributions downstream genome-wide, consistent with the hypothesis for these alleles having increased processivity and driving Pol II scanning further downstream than WT.

Predicted by our Shooting Gallery model, alleles with increased processivity would show expanded scanning windows, indeed, our downstream shifting *ssl2* alleles with potentially increased processivity show wider TSS distribution than the WT. In contrast, the Shooting Gallery model also predicts that alleles with decreased processivity would have narrower scanning windows than WT. Analysis of the TSS distribution in our upstream shifting *ssl2* alleles that have hypothetically reduced processivity showed narrower TSS window than the WT.

During transcription initiation, Pol II activity and Ssl2/TFIIH processivity both contribute to the scanning process. In this thesis, we showed that altered Pol II activity and TFIIH processivity shape the TSS distribution differently. For example, increased Pol II activity increases the efficiency of normally poorly used upstream TSSs, shifting TSS distributions upstream. Hypothetically, the more upstream space at a promoter, the more upstream TSS will be activated by increased Pol II activity. Consistent with this idea, we found that the degree of the upstream TSS shift is positively correlated with the distance between the PIC assembly site and the median TSS. In our examination of the relationship between the magnitude of the TSS shift and the PIC-TSS distance, we found that Pol II GOF alleles with increased catalytic activities shift more to the upstream site when the PIC-TSS distance is longer. In contrast, decreased Pol II activity cannot use upstream TSSs as efficiently as WT cells, resulting in apparently increased TSS usage at downstream sites. We hypothesized that, within the TFIIH defined processivity window, if there

is enough Pol II activity left from inefficient upstream initiation, the more downstream space is provided, the more Pol II LOF alleles can shift to the downstream sites. In other words, if TFIIH processivity were not limiting within a scanning window (*i.e*. there is remaining uninitiated Pol II when scanning is terminated by TFIIH processivity limits), the closer TSS is to the PIC, the greater chance that a Pol II LOF can still initiate downstream. Examination of the relationship between the degree of TSS shift in Pol II LOF allele and the PIC-TSS distance revealed that the shorter PIC-TSS distance is, the more Pol II LOF allele can shift to the downstream. We observed that the longer the PIC-TSS distance is, the more *ssl2* upstream shifting alleles can shift to the upstream sites, consistent with the observation that the TFIIH's processivity starts to decay at a relatively fixed position from where Pol II assembles. In addition, we found that the magnitude of the TSS shift in hypothetical *ssl2* processivity GOF allele is not sensitive the distance between the PIC and TSS. Indicating the degree of TSS shift in *ssl2* processivity GOF allele is determined by how much increased processivity can expand the downstream scanning window, which should have no difference at all promoter classes.

# 5.0 PRELIMINARY ASSESSMENT OF THE CONSERVATION OF THE SCANNING MECHANISM

## 5.1 INTRODUCTION

In *S. cerevisiae*, TSSs are distributed 40-120bp downstream of TATA-element for TATA-containing promoters; this TATA-TSSs distance is longer than in other eukaryotes. *Schizosaccharomyces pombe*, a fission yeast, uses TSS at 30-70nt downstream of the promoter that is conserved within metazoans. In addition, human TFIIB, which is involved in TSS selection, is compatible with the *S. pombe* transcription components *in vivo*[125]. Thus, *S. pombe* is a good and simple model to study TSSs selection more similar to metazoans. Our studies of Pol II active site mutants in *S. cerevisiae* support that promoter scanning is not only a mechanism for a few model genes, but for all types of promoters, suggesting that scanning is the universal mechanism for TSS selection in *S. cerevisiae*. We asked if scanning is also a mechanism in other eukaryotes. We first examined if scanning is a mechanism used in *S. pombe* by testing if the polar effect on TSS usage observed through perturbing Pol II's activity in *S. cerevisiae* would also be observed in *S. pombe*. A previous study has shown that the *S. cerevisiae* Pol II activity mutants' homologs in human also have altered elongation rates predicted by their rates in *S. cerevisiae*[126]. Therefore, we introduced mutations of the highly conserved trigger loop (TL) that shift TSSs in *S. cerevisiae* into the *S. pombe rpb1+* gene, which encodes the largest subunit of the pombe Pol II. The two mutants we created are *rpb1+* E1106G and *rpb1+* N1085S, which are analogous to *rpb1* E1103G and *rpb1* N1082S that confer higher and lower transcription activities in *S. cerevisiae* respectively. We next detected TSSs in these two mutants at selected model genes by primer extension and genome-wide

by TSS-seq. Though behaving differently from *S. cerevisiae* Pol II alleles, we found that *rpb1*+ N1085S showed effects on TSS usage both at model gene promoters and genome-wide. In contrast, *rpb1*+ E1106G exhibited fewer effects on TSS usage at tested promoters, however, TSS effects were observed at other promoters genome-wide. These observations together suggest a more complicated or different behavior of Pol II alleles in *S. pombe* TSS selection.

## 5.2 MATERIALS AND METHODS

### 5.2.1 *S .pombe* strains

*S. pombe* strains used in this chapter are listed in **Table 14**.

**Table 14 Yeast strains used in Chapter V**

| Strain number | Genotype |
|---|---|
| CKP001 | *h+ ura4-D18* |
| CKP010 | *h+ ura4-D18 rpb1*+ E1106G |
| CKP011 | *h+ ura4-D18 rpb1*+ E1106G |
| CKP031 | *h+ ura4-D18 rpb1*+ N1085S |
| CKP032 | *h+ ura4-D18 rpb1*+ N1085S |

### 5.2.2 *S. pombe* media

*S. pombe* medium used in this study were made as described in[127]. A brief summary, YE media is made of yeast extract (0.5% w/v; BD), glucose (3% w/v; BD). YEL is YE liquid media. YEA is YE media supplemented with bacto agar (2% w/v; BD). YES is YE media supplemented with adenine-HCl (1.31 mM, Sigma-Aldrich), L-histidine (1.45 mM, Sigma-Aldrich), L-leucine

(1.71 mM, Sigma-Aldrich), uracil (2.01 mM, Sigma-Aldrich) and L-lysine-HCl (1.23 mM, Sigma-Aldrich). Solid YES is additionally supplemented with bacto agar (2% w/v; BD). EMM media is made of EMM-glucose from Sunrise Science Products (CATALOG #: 2020-500) and 2% glucose. EMM+ media is EMM media supplemented with required amino acid at the above indicated concentrations.

### 5.2.3 Pop-in, pop-out allele replacement

To obtain *S. pombe* strains containing TL mutations, we applied a pop-in, pop-out allele replacement approach as previously described with slight modification[128]. Briefly, the *rpb1+* gene containing wanted mutations (*) were cloned into a *pUC8* plasmid that harbors a *ura4+* gene (*pUC8-ura4+*). This plasmid was then enzymaticly linearized and transformed into a 1.8kb genomic *ura4* locus deleted (*ura4-D18*) fission yeast strain using a lithium acetate protocol described in[127]. The transformed cells were plated on EMM media supplemented with other required amino acids but uracil (EMM+Ade+His+Leu+Lys), and cultured at 32˚C for 4 days. The plasmid *rpb1+* sequence flanking *ura4+* is homologous to the endogenous *rpb1+* sequence and allows for homologous recombination between the two (pop-in). Successful recombination will integrate the plasmid *rpb1+ ura4+* genes into the genomic *rpb1+* gene and confer a $Ura^+$ phenotype. $Ura^+$ colonies growing on the EMM+ selection media were streaked on the same media to ensure the selection fidelity. Selected single colonies were then incubated in 10mL of YES liquid media at 32˚ C for 16-20h to final OD595 of 2-4. Cell numbers were counted as described to determine the OD[127]. Liquid cell culture was then plated on YEA medium containing 1mg/mL 5FOA and incubated at 32˚ C for 5-7 days to select for cells that have lost *ura4+* and are $Ura^-$

(putative pop-outs). 5-FOA-resistant colonies were streaked on one YEA+5FOA plate. Genomic DNA of 5-FOA-resistant colonies were extracted for PCR genotyping.

### 5.2.4 RNA extraction, Primer extension and TSS-seq

RNA extraction, primer extension and the TSS-seq were performed the same as described in 2.2.5 and 4.2.1 except for using YES as media to culture fission yeast cells.

### 5.2.5 *S. pombe* TSS-seq data analysis:

Data analysis for *S. pombe* TSS-seq is the basically the same as for *S. cerevisiae* TSS-seq data analysis except for minor difference in mapping (see below 5.2.6.1) and making TSS count table (see below 5.2.6.2).

#### 5.2.5.1 Mapping

Mapping was performed as for *S. cerevisiae* described in 4.2.2.1 except for using pombe genome ASM294v2 (EF2) as a reference (downloaded from https://support.illumina.com/sequencing/sequencing_software/igenome.html).

#### 5.2.5.2 TSS count table and the heatmap

For 4021 pombe promoters, the 5' ends of the mapped sequencing reads within each of the previously annotated 101 bp TSS window[129], spanning 50 nucleotides upstream and 50 nucleotides downstream of the annotated WT median TSS, was extracted from the bedgraph file using deepTools to make a 101×4041 TSS count table[130]. Each row of the TSS count table represents a

promoter position, and the number in each cell representing 5' ends mapped to that position. The count tables generated from two biological replicates were merged into one by aggregating read counts from two. The merged count table was row-wise normalized to get the relative TSS usage in each promoter, rank ordered by promoter total reads and visualized by Morpheus heatmap. TSS distribution differences were determined by subtracting normalized WT data from normalized mutant data and visualized in Morpheus heatmap.

## 5.3 RESULTS

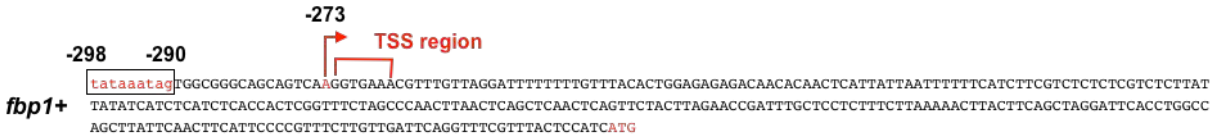
### 5.3.1 TSS usage of Pol II mutants with potentially altered catalytic activity in fission yeast *S. pombe*

To examine if *S. pombe* TSSs are similarly sensitive to putative perturbation of Pol II activity in fashion predicted from the scanning model, we created Pol II mutants in *S. pombe* that potentially have altered catalytic activity and tested their TSS usage at selected gene promoters and genome-wide. These two Pol II mutants are *rpb1*+ E1106G and *rpb1*+N1085S, which contain amino acid substitutions that are homologous to *S. cerevisiae* E1103G and N1082S. They are hypothesized to have altered catalytic activities similar to their homologs in *S. cerevisiae* due to the high conservation of the TL sequence, structure and the catalytic mechanisms in Pol II transcription.

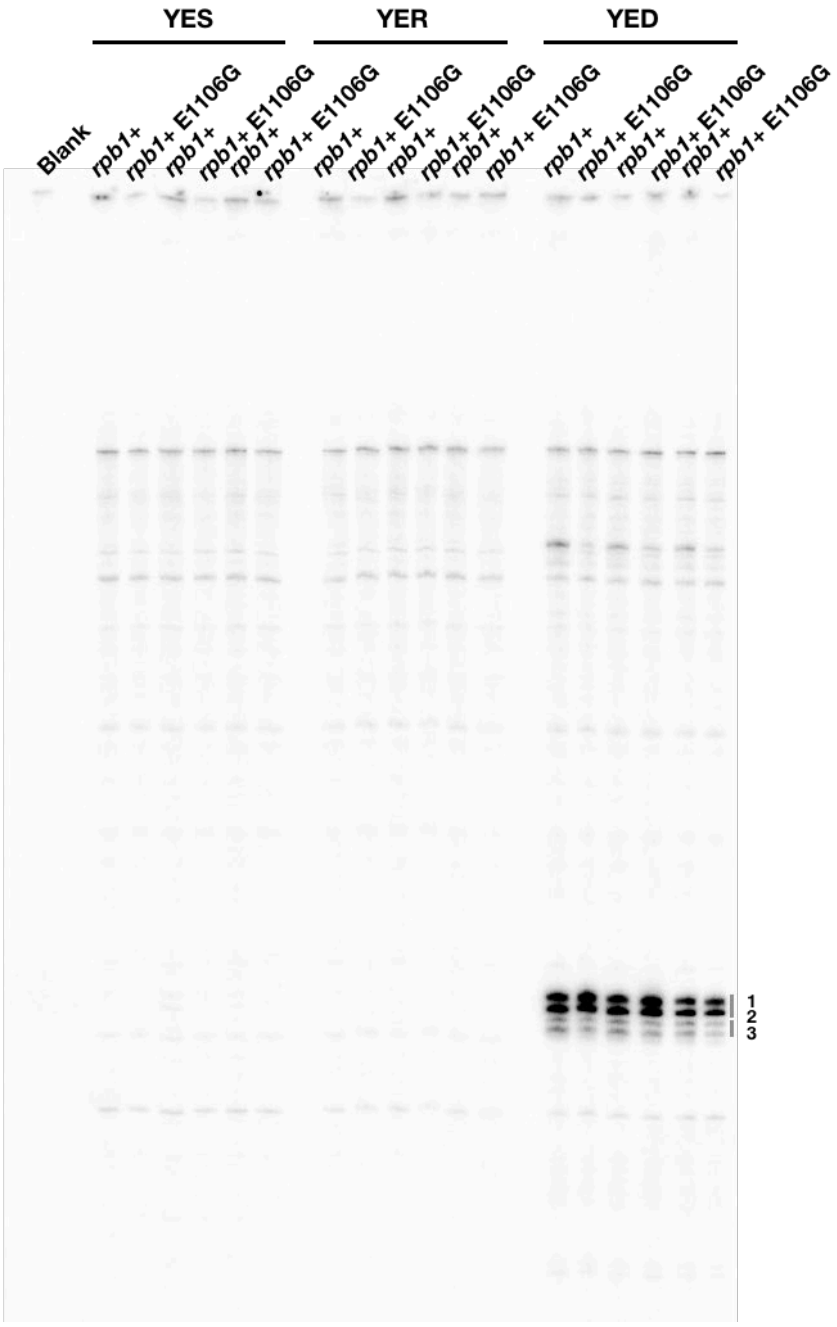
We first detected TSS usage by primer extension at a model gene *fbp1*+, which had been previously analyzed by Hoffman *et. al*[131] **(Figure 44A)**. This gene contains multiple TSSs that were previously annotated 26-33 bp downstream of the putative TATA-box and its expression is

induced under glucose starvation. Therefore, glucose deficient medium YPD was used to induce *fbp1+* expression, the normal yeast media YES and glucose rich median YER are used as control **(Figure 44B)**. We observed that after 60 minutes of glucose starvation, TSSs are activated. Visually, *rpb1+* E1106G allele's effect on TSS usage at *fbp1+* in *S. pombe* is not strong as *rpb1* E1103G's effect on TSS usage at *ADH1* in *S. cerevisiae*. However, quantification of TSS signal showed that, compared to the WT, TSS usage of *rpb1+* E1106G in bin 1 is increased, whereas TSS usage in bin 3 is decreased **(Figure 44C)**.

**A**


-273
TSS region
-298
-290
fbp1+
tataaatagTGGCGGGCAGCAGTCAAGGTGAAACGTTTGTTAGGATTTTTTTTGTTTACACTGGAGAGAGACAACACAACTCATTATTAATTTTTTCATCTTCGTCTCTCTCGTCTCTTAT
TATATCATCTCATCTCACCACTCGGTTTCTAGCCCAACTTAACTCAGCTCAACTCAGTTCTACTTAGAACCGATTTGCTCCTCTTTCTTAAAAACTTACTTCAGCTAGGATTCACCTGGCC
AGCTTATTCAACTTCATTCCCCGTTTCTTGTTGATTCAGGTTTCGTTTACTCCATCATG

**B**


YES
YER
YED
Blank
rpb1+
rpb1+ E1106G
1
2
3

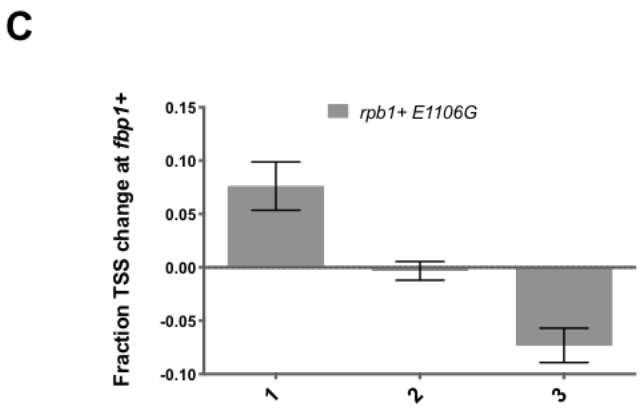



**Figure 44 TSS usage in *rpb1*+ E1106G at *fpb1*+ promoter uder glucose starvation**

**(A)** Schematic showing of *fbp1*+ promoter architecture. The *fbp1*+promoter contains a nominant TSS that is 26 bp (-273 relative to start codon) downstream of the putative TATA-box, followed by a few other TSS 7 bp further downstream of the dominant TSS (other TSS is not well charicterized but indicated previously)[131]. **(B)** *rpb1*+ E1106G affects TSS usage at *fbp1*+ promoter. Primer extension shows that, 60 mininutes after the glucose starvation, *fbp1*+ gene's expression is induced, we observed that, compared to the WT, *rpb1*+ E1106G uses more dominant TSS and less TSS downstream sites. **(C)** *rpb1*+ E1106G quantitatively increase TSS usage upstream and decrease TSSs usage downstream. TSS signal at *fbp1*+ promoter is divided into three bins. As previously introduced, change of TSS usage at each bin is measured by subtracting the calculated percentage of WT signal from that of the mutant strain. Compared to the WT, TSS usage in bin 1 is increased, whereas TSS signal in bin 3 is decreased.

To test *rpb1*+ alleles' effect on TSS usage on other gene promoters, we selected seven highly expressed genes in *S. pombe* and examined *rpb1*+ E1106G allele's effect on these promoters' TSS usage by primer extension. These promoters are listed in **Table 15**. Some of these promoters contain TATA consensus sequence "TATAAAA", *e.g. rps802* and *cbr1* have TATA elements 386 bp and 234 bp upstream of their start codons respectively. *hxk2* and *ubi1* also contain TATA elements 951 bp and 719 bp upstream of their correspondent start codons, however, they are too far away from the start codon and less likely to be the CPE for these two promoters. Other

selected promoters don't contain featured TATA elements. We next tested TSS usage of these seven promoters using only one biological replicate to get a broad idea for their behavior in our primer extension system. Our primer extension results showed that there is no clear TSS signal detected at the promoter region of *cbr1* we were priming **(Figure 45)**. In addition, *wis1* and *hxk2* show very weak TSS signal. We therefore decided to use promoters SPBC725.01, *rps802, pfk1* and *ubi1* for downstream analysis.

**Table 15 Promoters selected for *S. pombe* primer extension**

| Gene name | Alias | TATA containing or not |
|---|---|---|
| ***wis1*** | SPBC409.07c | No TATA |
| **Unassigned** | SPBC725.01 | No TATA |
| ***rps802*** | SPAC521.05 | 386 upstream of start codon TATA |
| ***cbr1*** | SPCC970.03 | 234 upstream of start codon TATA |
| ***pfk1*** | SPBC16H5.02 | No TATA |
| ***hxk2*** | SPAC4F8.07c | 951 upstream of start codon TATA |
| ***ubi1*** | SPAC11G7.04 | 719 upstream of start codon TATA |

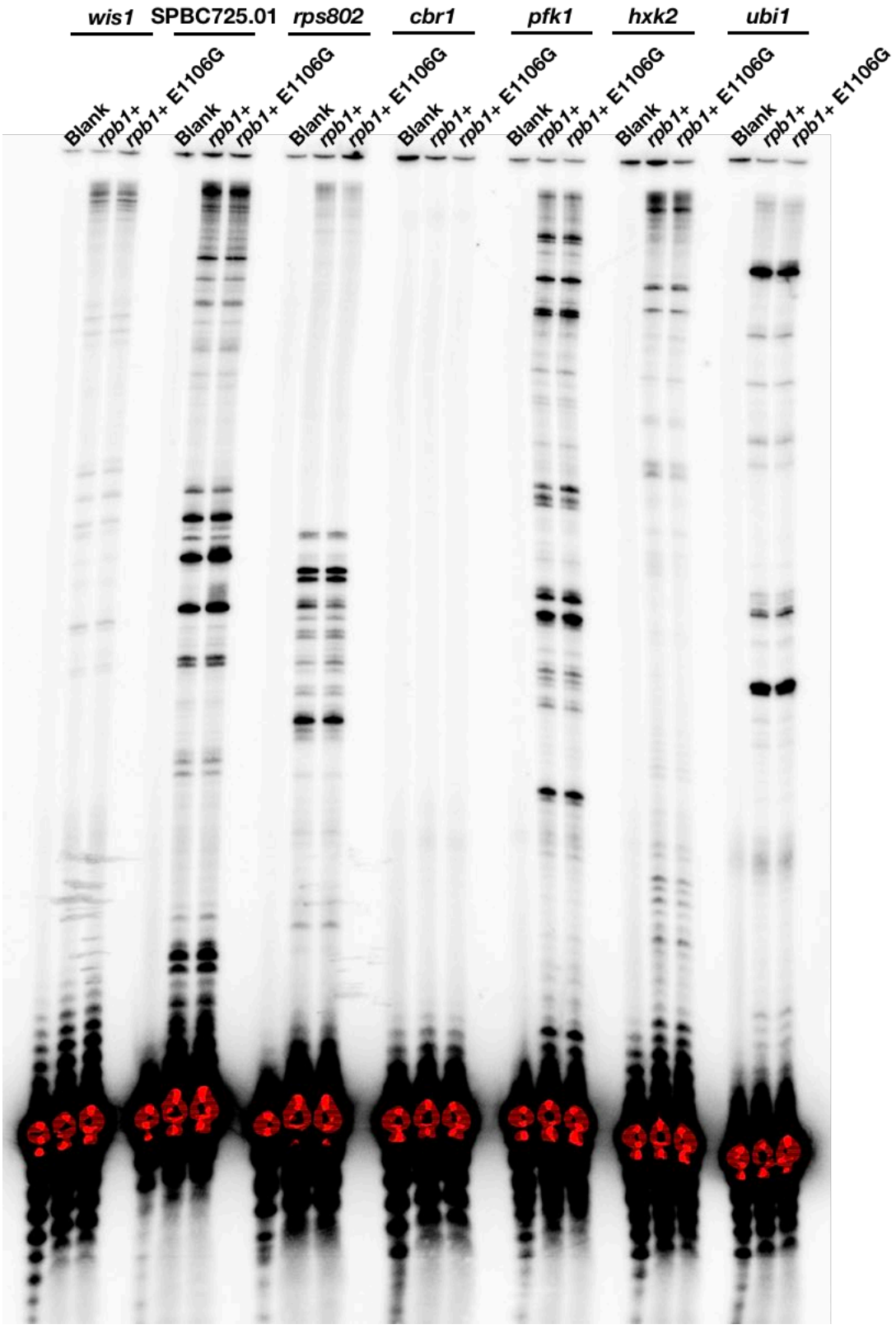
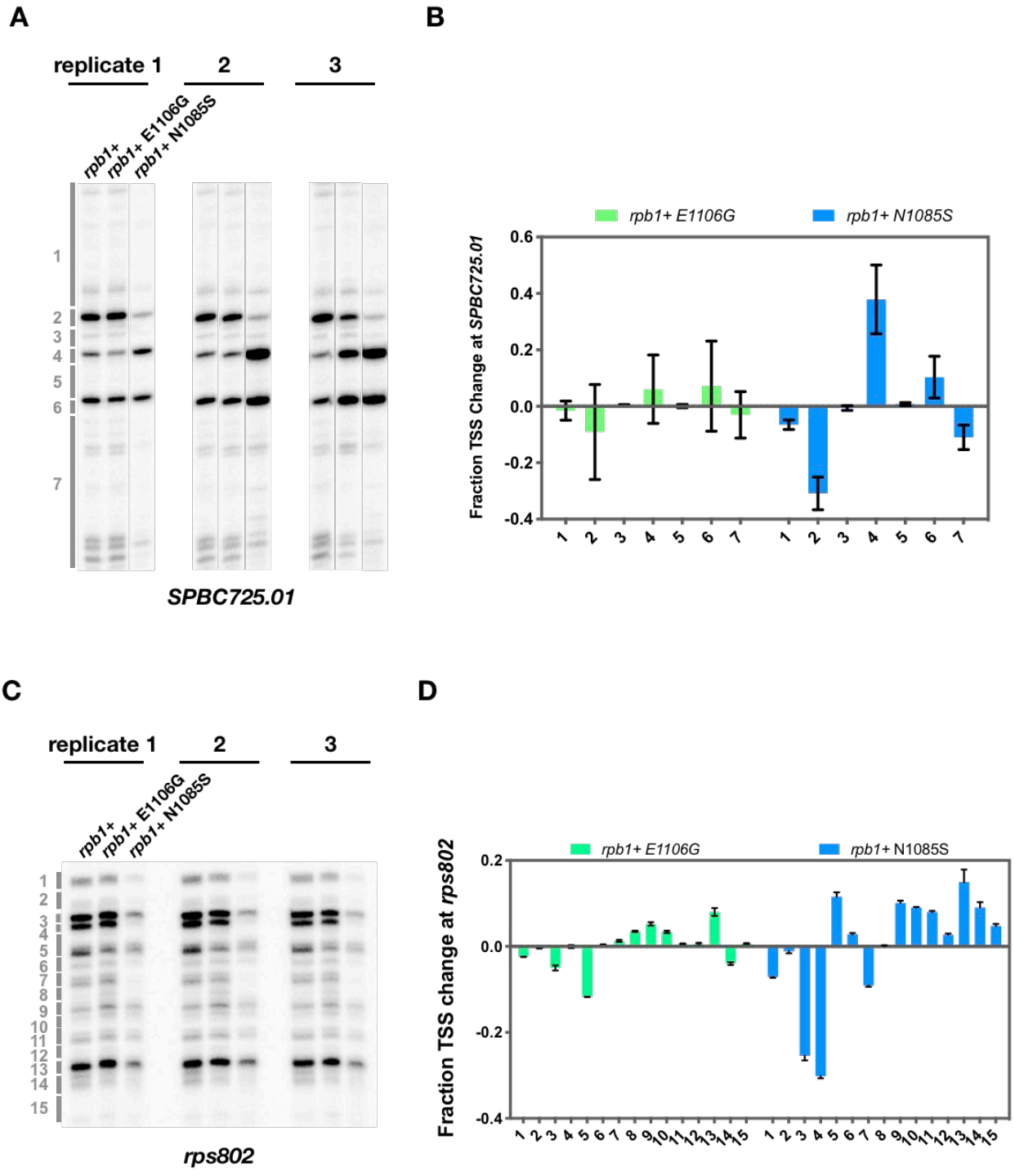



**Figure 45 TSS usage at seven selected *S. pombe* promoters**

Primer extension results show that four (SPBC725.01, *rps802*, *pfk1* and *ubi1*) out of the seven selected promoters show TSS signal at the priming regions. *cbr1* show no clear TSS signal, *wis1* and *hxk2* show very weak TSS signal at tested conditions.

Next, we performed primer extension to test *rpb1*+ E1106G and *rpb1*+ N1085S allele's effect on TSS usage for both *rpb1*+ E1106G and *rpb1*+ N1085S at promoters of SPBC725.01, *rps802*, *pfk1* and *ubi1*. **(1)** Our primer extension results showed that allele E1106G had little effect on TSS usage at SPBC725.01 **(Figure 46A)**. Similar to what we have showed on quantification of TSS usage at *ADH1* promoter, we divided TSS signal at SPBC725.01 promoter region into 7 bins and, compared to the WT, only slight change was observed in bin 2, 4 and 6 in E1106G allele **(Figure 46B)**. In contrast, N1085S showed dramatic effect on TSS usage at SPBC725.01, it decreased TSS usage at one of the three upstream major site (bin 2) and increased TSS usage at two downstream major sites (bin 4 and 6) **(Figure 46A)**. N1085S allele's effect was quantitatively showed as ~30% decreased usage at upstream bin 2 and ~40% increased usage at downstream bin 4 **(Figure 46B)**. The direction of *S. pombe* allele N1085S in shifting TSS usage is similar to what we have observed in its *S. cerevisiae* homolog N1082S allele, both shifting TSS usage downstream. However, different from the behavior of *S. cerevisiae* homolog N1082S in shifting TSS downstream through increasing the apparent TSS efficiency at normally poorly used downstream sites, *S. pombe* allele N1085S didn't show increased activity at downstream minor sites. There are a few possibilities for this difference, first, the promoter architecture could be different between *ADH1* and SPBC725.01, for example, the existence of the downstream constrains at SPBC725.01 promoter for limiting N1085S to activate downstream minor sites. Alternatively, the apparent TSS shifting difference between *S. pombe* and *S. cerevisiae* indicates that *S. pombe* could have used a different mechanism for selecting TSS rather than a scanning

mechanism. **(2)** Allele E1106G exhibited little effect on TSS usage at *rps802* promoter, exhibiting a slight decreased on some of the upstream sites (bin 1, 3 and 5) and minor increase at some of the downstream sites (bin 8-11 and 13) **(Figure 46C, 46D)**. Strikingly, allele N1085S decreased the overall TSS usage at *rps802* promoter, showing as dramatically decreased TSS signal on the gel **(Figure 46C)**. Interestingly, quantification of TSS usage of N1085S showed a clear downstream shift at *rps802* promoter compared to the WT, consistent with *S. cerevisiae* N1082S allele's behavior in shifting TSS distribution downstream **(Figure 46D)**. Again, *S. pombe* N1085S didn't increase the apparent downstream minor TSS usage. **(3)** Similar to its behavior at SPBC725.01 and *rps802* promoters, *S. pombe rpb1+* E1106G showed almost no effect on TSS usage at *pfk1* and *ubi1* promoters compared to the WT **(Figure 46E-H)**. Although *rpb1+* N1085S didn't shift TSS distribution in a polar fashion at *pfk1* and *ubi1* promoters, the TSS usage pattern is clearly different than in the WT. For example, N1085S shifted TSS usage downstream in a small region of *pfk1* promoter (from bin 1 to bin 2), and in a small region of *ubi1* promoter (bin 2). To summarize, *S. pombe rpb1+* E1106G had little effect on TSS usage at the above four tested promoters, whereas *rpb1+* N1085S changed TSS usage at all four tested *S. pombe* promoters. The way that *S. pombe rpb1+* N1085S allele changes TSS usage showed some similarity to its *S. cerevisiae* homolog N1082S, both shifting TSS downstream, but distinct in the absence of apparent increased usage of downstream minor TSSs.


A
replicate 1
2
3
rpb1+
rpb1+ E1106G
rpb1+ N1085S
1
2
3
4
5
6
7
SPBC725.01
B
rpb1+ E1106G
rpb1+ N1085S
Fraction TSS Change at SPBC725.01
0.6
0.4
0.2
0.0
-0.2
-0.4
1 2 3 4 5 6 7 1 2 3 4 5 6 7
C
replicate 1
2
3
rpb1+
rpb1+ E1106G
rpb1+ N1085S
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
rps802
D
rpb1+ E1106G
rpb1+ N1085S
Fraction TSS change at rps802
0.2
0.0
-0.2
-0.4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

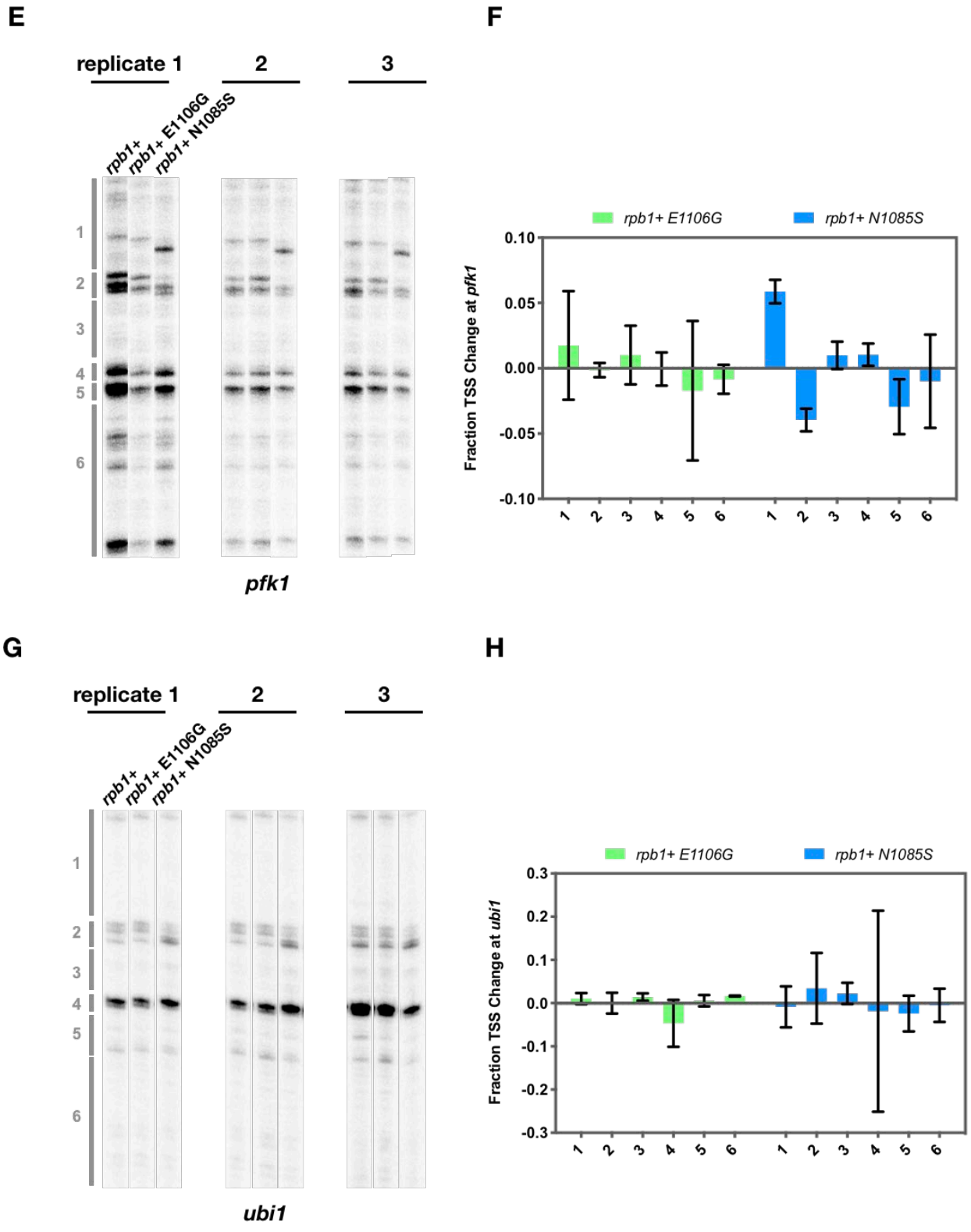



**Figure 46 *S. pombe* Pol II alleles showed both similarities and distinctions in TSS usage to their *S. pombe* homologous alleles**

**(A)** *S. pombe rpb1*+ N1085S shifts TSS downstream at SPBC725.01 promoter, wherase E1106G showed little effect on TSS usage. Primer extension results showed no clear difference for TSS usage in E1106G relative to the WT. E1106G increased TSS usage at downstream sites of SPBC725.01 promoter (TSS in bin 4 and 6). **(B)** Quantification of TSS usage at SPBC725.01 promoter in *S. pombe rpb1*+ N1085S and E1106G. **(C)** *S. pombe rpb1*+ N1085S reduced the overall TSS usage at *rps802* promoter, wherease E1106G showed little effect. Primer extenstion results showed that the TSS usage in E1106G was not changed comapred to the WT, in contrast, N1085S reduced the overall TSS signal at *rps802* promoter. **(D)** Quantification of TSS usage at *rps802* promoter in *S. pombe rpb1*+ N1085S and E1106G. N1085S quantitatively increased TSS usage at downstream sites and increased relative TSS usage upstream sites. E1106G showed little increase at dowsntream sites. **(E)** *S. pombe rpb1*+ N1085S showed minor change of TSS usage at *pfk1*. Primer extension results showed slight TSS usage difference between the WT and N1085S at the upstream region of *pfk1* promoter (bin 1 and 2). E1106G showed almost no effect on TSS usage at *pfk1*. **(F)** Quantification of TSS usage at *pfk1* promoter in *S. pombe rpb1*+ N1085S and E1106G. **(G)** *S. pombe rpb1*+ N1085S showed slight change of TSS usage at *ubi1*. Primer extension results showed slight TSS usage difference between the WT and N1085S at the upstream region of *ubi1* promoter (bin 2). E1106G showed almost no effect on TSS usage at *ubi1*. **(H)** Quantification of TSS usage at *ubi1* promoter in *S. pombe rpb1*+ N1085S and E1106G.

### 5.3.2 Genome wide impact of Pol II mutants in TSS usage in fission yeast *S. pombe*

Mapping statistics of the *S. pombe* TSS-seq shows that, similar to data analysis in *S. cerevisiae*, around 50-70% RNA reads are aligned to the pombe genome **(Table 16)**. Among the other 30-50% unaligned reads, 10% of which are *S. cerevisiae* RNAs previously added as a spike-in control.

**Table 16 *S. pombe* TSS-seq mapping statistics**

| Sample ID (Vv) | CKP | Mutant | Number of raw reads | Reads with at least one | Reads failing to align | Reads aligned more than |
|---|---|---|---|---|---|---|

| | | | | **reported alignment** | | **once** |
|---|---|---|---|---|---|---|
| 1390 | CKP001 | *rpb1*+ WT-1 | 24,950,519 | 17954249 (71.96%) | 3136514 (12.57%) | 3859756 (15.47%) |
| 1391 | CKP001 | *rpb1*+ WT-2 | 33,015,409 | 23009220 (69.69%) | 4011288 (12.15%) | 5994901 (18.16%) |
| 1392 | CKP021 | *rpb1*+ E1106G-1 | 37,085,088 | 23810537 (64.21%) | 3986288 (10.75%) | 9288263 (25.05%) |
| 1393 | CKP022 | *rpb1*+ E1106G-2 | 37,051,561 | 26155282 (70.59%) | 4218489 (11.39%) | 6677790 (18.02%) |
| 1423 | CKP031 | *rpb1*+ N1085S-1 | 34,642,157 | 19344815 (55.84%) | 6317753 (18.24%) | 8979589 (25.92%) |
| 1424 | CKP032 | *rpb1*+ N1085S-2 | 27,207,384 | 13932629 (51.21%) | 4771984 (17.54%) | 8502771 (31.25%) |

The positions and counts of the 5' ends of the uniquely mapped reads from *S. pombe* TSS-seq data were extracted using deepTools to make a 5' ends count table as described in Materials and Methods. Information stored in the count table was then used to assess the reproducibility between biological replicates through calculating the Pearson correlation coefficient and the correlation map between replicates. We found that all the replicates in *S. pombe* TSS-seq showed higher than 0.9 Pearson correlation coefficient **(Figure 47)**.

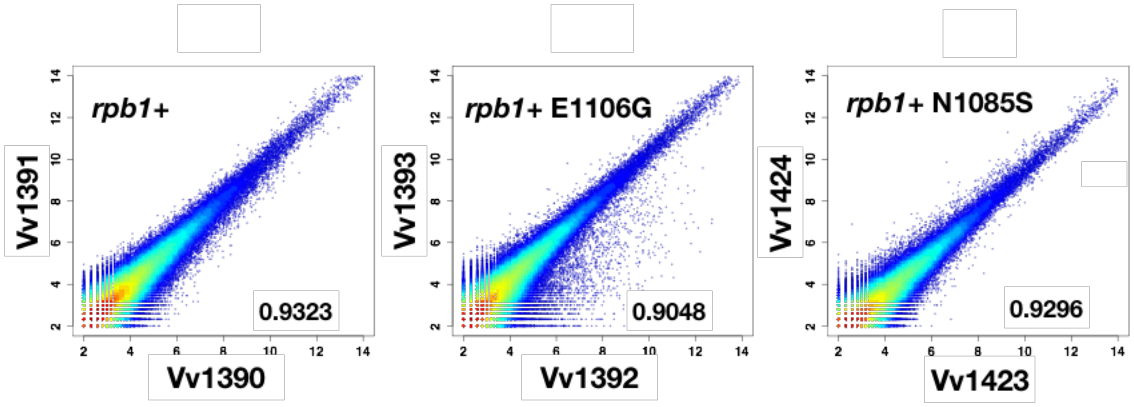



**Figure 47 Correlation of reads number between *S. pombe* replicates at single TSS level**

Scatter plot showing the correlation of the reads number at individual TSS between *S. pombe* TSS-seq replicates. All the replicates in S. pombe TSS-seq showed great correaltion as indicated by the high correlation coefficient number and the focused correlation plot.

Hierarchical clustering on Pearson correlation coefficients among the *S. pombe* TSS-seq libraries showed that the replicates of the same WT/mutant are closely clustered and the three mutants are distinguished into 2 large groups (I and II) or 3 subgroups (I, IIa and IIb) **(Figure 48)**. In hierarchical clustering, mutant *rpb1*+ N2085S distinguishes itself from WT and *rpb1*+ E1106G by having a greater Euclidean distance than distance between the WT and *rpb1*+ E1106G. This is consistent with N2085S allele's behavior in making changes of TSS usage at tested model genes, whereas WT and E1106G are more similar in TSS usage **(Figure 46)**.

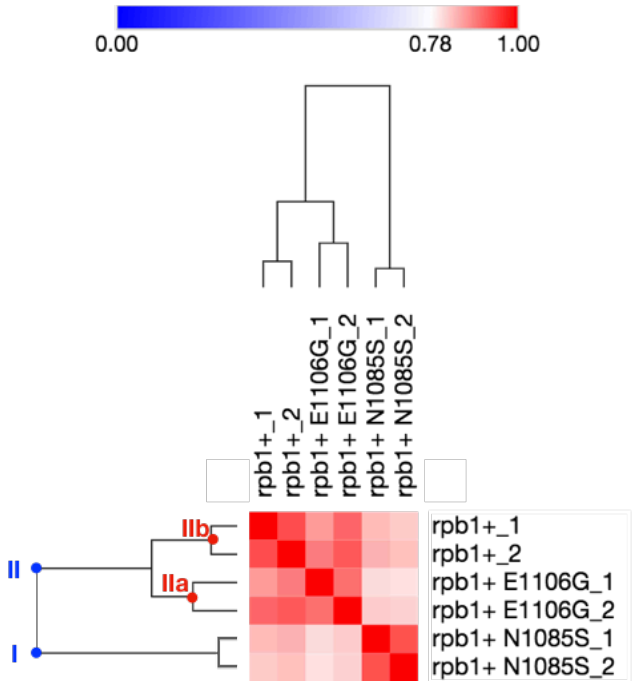



**Figure 48 Matrix heatmap and the hierarchical clustering of Pearson correlation coefficients between *S. pombe* TSS-seq libraries**

Matrix showing the Pearson correlation coefficients among *S. pombe* TSS-seq libraties. *Rpb1* TSS-seq ibraries are clusterd into two large groups I and II, or into three subgroups 1-3. *rpb1*+ N1085S is distinguished from the WT and *rpb1*+ E1106G, consisting with its behavior in changing the TSS usage at models genes, whereas E1106G and WT showed more similarity in TSS usage at model genes.

TSS distribution of 4041 *S. pombe* promoters showed that *S. pombe rpb1+* alleles don't change TSS distribution as dramatically as *S. cerevisiae rpb1* alleles did **(Figure 49A)**. TSS distribution in *rpb1+* E1106G, H1085S and the WT strains show no great difference. However, changes of TSS usage in TSS difference maps indicated that both alleles shift TSS usage at most of the pombe gene promoters, showing as increased TSS signal (in orange) upstream of the promoters in TSS difference map between *rpb1+* E1106G and WT, and decreased TSS usage (in cyan) at upstream of most promoters in the TSS difference map between *rpb1+* H1085S and the WT **(Figure 49B)**. According to the observations of TSS usage at model genes, genome-wide TSS shifting in *rpb1+* H1085S alleles could be either caused by decreased TSS usage at upstream of sites and increase TSS usage at downstream sites, like in SPBC725.01 promoter, or resulted from decreased TSS usage at upstream sites but no activation at downstream sites, like in *rps802* promoter.

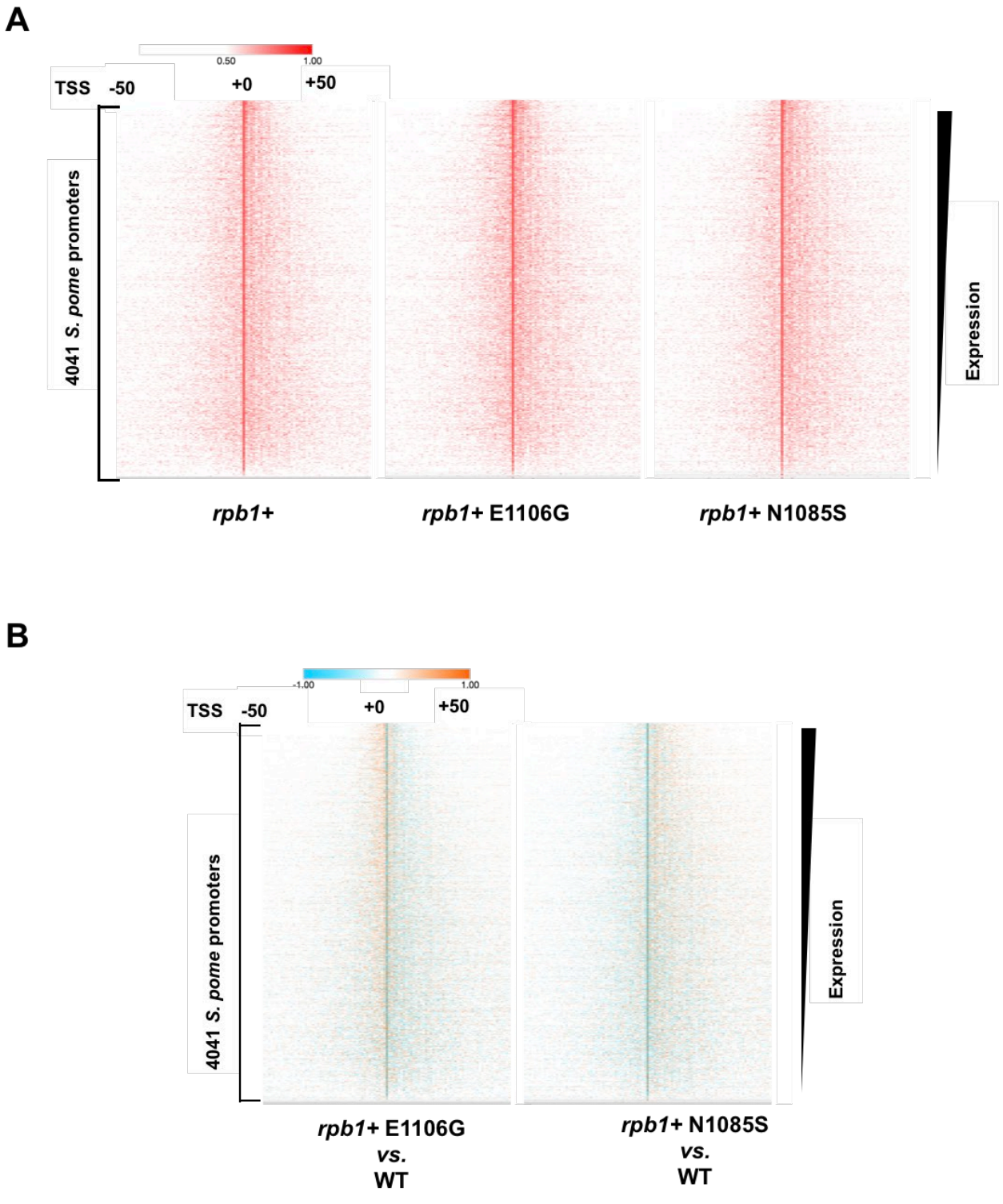



**Figure 49 *rpb1+* mutants alter TSS ditribution genome-wide**

**(A)** TSS distribution of *rpb1+* WT, E1106G and N1085S. Heatmap showing TSS read distributions at WT median TSS centered 4041 pombe promoter windows. Promoters are rank ordered according to the expression level (total

reads) in the WT strain. TSS read distributions show no clear difference between the WT and E1106G. In contrast, compared to the WT, there is slight TSS increase at downstream sites of N1085S. **(B)** *rpb1+* alleles alter TSS distribution at most of genome promoters. Heatmap of TSS distribution difference between WT and mutants at median TSS centered window shows that, compared to the WT, E1106G decreases TSS signal at downstream region and increases TSS signal at upstream sites at most of genome promoters. In contrast, N1085S show oppisite changes, downstream TSS usage is increased and the relative upstream TSS signal is decreased.

The effect of *rpb1+* allele on TSS usage was quantified by measuring the median TSS shift between *rpb1+* mutants and WT. Though, more promoters appear to show upstream median TSS shift than the downstream median TSS shift genome-wide in *rpb1+* E1106G, **(Figure 50, more data points below the x-axis than above x-axis)**. The positions of the median TSS between *rpb1+* E1106G and WT showed no significant statistical difference **(Figure 50)**. In contrast, the positions of the median TSS between *rpb1+* N1085S and WT showed significant difference. The direction of the TSS shift in *rpb1+* N1085S is the same as observed at primer extension tested model genes and in the differential TSS distribution map, showing as 1bp of 75% percentile downstream shift. Overall, the degree of TSS shift in *S. pombe rpb1+* alleles is not as significant as observed in *S. cerevisiae rpb1* alleles.

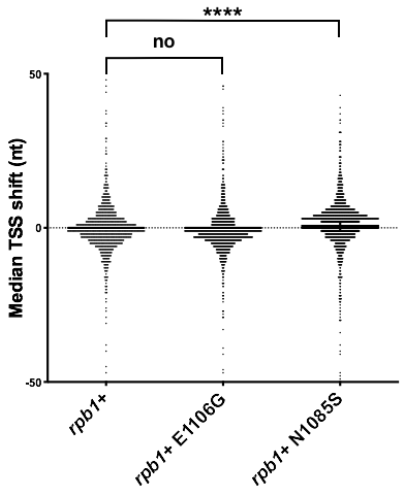



**Figure 50 The degree of TSS shift in *S. pombe rpb1+* mutants is not as significant as TSS shift in *S. cerevisiae rpb1* mutants**

Boxplot showing the median TSS shift between WT and *rbp1+* mutants. One-way ANOVA test of the median TSS shift between WT and *rpb1+* alleles shows that there is no significant difference between WT and *rpb1+* E1106G, wherease there is a significant difference between WT and *rpb1+* N1085S.

Examination of the relationship between the median TSS shift and promoter expression level of *rpb+* mutants showed that there is no significant correlation between the two in both mutants **(Figure 51)**. Interestingly, by showing median TSS shift in scatter plot, we observed a clear distribution of the majority of median TSS shifts in *rpb+* E1106G below the x-axis, indicating TSS upstream shifting. In contrast, most the median TSS shift in *rpb+* N1085S were above the x-axis, suggesting TSS downstream shifting.

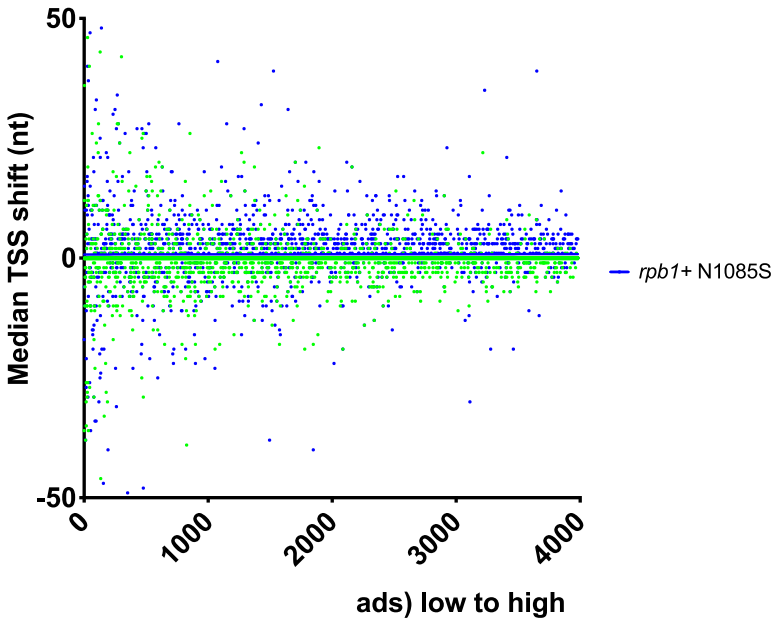



**Figure 51 The relationship between the median TSS shift and the promoter expression level**

Scatter plot showing the relationship between the median TSS shift and the expression level in WT and *rpb1+* mutants. There is no clear trend observed between the two in both alleles.

## 5.4 DISCUSSION

Studies have shown that *S. cerevisiae* uses the scanning mechanism for TSS selection during Pol II transcription initiation. Whether or not scanning a conserved mechanism for Pol II TSS selection among eukaryotes is an unanswered question. Here, we examined TSS usage in *S. pombe*, a fission yeast that has conserved TATA-TSS distances with higher eukaryotes, to examine if the scanning mechanism is conserved between *S. cerevisiae* and *S. pombe*. We used *S. pombe* alleles containing mutations within the highly conserved Pol II TL domain that hypothetically alter Pol II catalytic activity and tested TSS usage at a few model genes and genome-wide. We have found that *S. pombe* Pol II alleles with putatively decreased catalytic activity shift TSS

distributions downstream both at tested model genes and genome-wide. However, the pattern of TSS distribution shifts in *S. pombe* is different from how Pol II catalytic LOF alleles shift TSS distributions in *S. cerevisiae*. In *S. cerevisiae*, Pol II LOF alleles shift TSS downstream by extending initiation to minor TSSs downstream, whereas *S. pombe* shifts TSS usage between major TSSs. In addition, the putative *S. pombe* GOF allele didn't show strong effects on TSS usage at tested promoters; however, an upstream shifting trend was broadly observed genome-wide by comparing the median TSS positions between the WT and the mutant. We interpret this difference for TSS usage between *S. pombe* and *S. cerevisiae* as follows: first, there is a possibility that not all the pombe promoters use a long (20-100 nt) scanning mechanism for TSS selection but instead can scan in a highly localized window; second, this difference may be due to different promoter architectures between budding and fission yeast. For example, different upstream and downstream constraints for limiting TSS shifts; third, the actual catalytic activities for *S. pombe* alleles used in this study haven't been tested, and there is a possibility that the catalytic activity is not as strongly altered as in *S. cerevisiae*, resulting in smaller perturbations to putative scanning.

# 6.0 CONCLUSIONS AND FUTURE DIRECTIONS

## 6.1 CONCLUSIONS

This thesis studied how Pol II and Ssl2 function in transcription initiation and TSS selection. The work described in this thesis mainly contains two parts. First, utilizing genetic tools and genomic approaches, the function of Ssl2 in TSS scanning was dissected in the model organism budding yeast *S. cerevisiae*. Second, potential conservation of initiation by scanning, which is used by budding yeast for TSS selection, was investigated in two additional model organisms, the fission yeast *S. pombe* and the fruit fly *D. melanogaster* (see Appendix A).

Our studies of five pre-existing *ssl2* mutants at budding yeast promoters showed transcription-related growth phenotypes, suggesting defects in transcription initiation and TSS selection. Primer extension of TSS usage in some of these mutants revealed polar shifts in TSS usage at the model gene *ADH1*. We observed that *ssl2* allele TSS distribution shifts were qualitatively and quantitatively different from how Pol II alleles and other transcription factor alleles, *e.g.* TFIIF alleles, shift TSS distribution at *ADH1*. For example, we observed that some *ssl2* alleles shift TSS distribution upstream through limiting TSS usage at promoter downstream sites, whereas Pol II catalytic activity GOF alleles with increased elongation rate shift TSS distribution upstream by activating normally poorly used upstream TSSs. *ssl2* upstream-shifting alleles' defects in downstream sites were consistent with a recent study, which showed how Ssl2 promotes downstream DNA opening during promoter scanning[44].

We hypothesize that Ssl2 functions in driving Pol II scanning during transcription initiation and TSS selection. Specifically, it determines how far (processivity) and how fast (translocation

rate) Pol II machinery can go downstream. We interpret the coordination of Ssl2 activity in driving transcription machinery scanning and Pol II activity in activating TSS usage coordinated during promoter scanning as part of a "Shooting Gallery" model. The Shooting Gallery model predicts that Ssl2's processivity determines how wide the scanning region is and therefore shapes the widths of TSS distributions. It also predicts that Pol II activity affects TSS distribution through changing the efficiency within Ssl2's processivity-defined scanning window.

To test the Shooting Gallery model, we examined the genetic interactions between Pol II and *ssl2* mutants. We started by extending our *ssl2* mutant pool by genetic screening. Our genetic screen identified 42 *ssl2* single mutants that were defective in TSS selection. These *ssl2* mutants fell into two major classes in terms of TSS distribution at *ADH1* promoter and *IMD2*-related transcription phenotypes. One class shifted TSS distribution downstream at *ADH1*, exhibited $His^+$ phenotype and hypothetically increased Ssl2 processivity. The second class shifted TSS distribution upstream at *ADH1,* showed $MPA^S$ phenotypes and hypothetically reduced Ssl2 processivity. Among the identified *ssl2* downstream shifting alleles, a small group additionally exhibited a $Spt^-$ phenotype that had not been identified previously in other *ssl2* mutants. Mutations that confer $Spt^-$ phenotypes were clustered in the N-lock region of Ssl2 and could represent a new functional class. The genetic interactions between Pol II/GTF and Ssl2 alleles supported the idea of two major networks controlling TSS selection, with one network shaping TSS distribution through affecting initiation efficiency. This network is represented by Pol II, TFIIB and TFIIF. A second network hypothetically affects TSS distribution through regulating TFIIH's processivity, and includes Ssl2 and Sub1.

Our genome-wide study of Ssl2 allele effects on TSS usage revealed that TSS distributions were altered for the majority of promoters and for all promoter classes we examined. *ssl2* alleles

that exhibited MPA$^S$ phenotypes and shifted TSS upstream at *ADH1* decrease downstream TSS usage shifted TSS upstream genome-wide. In contrast, *ssl2* alleles showed His$^+$ phenotypes and shifted TSSs downstream at *ADH1* also shifted TSS distributions downstream genome-wide. *ssl2* alleles exhibited lesser effects on TSS shift, which was measured by the degree of the median position change between WT and mutant, than Pol II alleles. In addition, consistent with the Shooting Gallery model's prediction, *ssl2* downstream shifting alleles, the hypothetical processivity GOF class, increased the widths of TSS distributions genome-wide. In contrast, *ssl2* upstream shifting alleles, the hypothetical processivity LOF class, decreased the widths of TSS distributions genome-wide. The relationship between the extent of the TSS shift and the distance between the PIC assembly site and the median TSS positions (PIC-TSS distance) showed different patterns in two classes of *ssl2* alleles. The degree of TSS shift in *ssl2* upstream shifting alleles was sensitive to the PIC-TSS distance: the further the PIC-TSS distance was, the greater the observed shift in *ssl2* upstream shifting alleles. These results were consistent with the idea that promoters with TSSs evolved at more downstream positions from the PIC assembly point would be predicted to be more likely limited by Ssl2 processivity, and therefore more sensitive to *ssl2* alleles. Additionally, promoter expression was also inversely correlated with PIC-TSS distance in *ssl2* upstream shifting alleles: the further downstream TSSs were, the greater the truncation of TSSs by *ssl2* alleles and the greater the reduction in promoter expression. In contrast, the magnitude of TSS shift in *ssl2* downstream shifting alleles was not sensitive to the PIC-TSS distance, potentially due to coupling of Pol II activity and TFIH processivity during promoter scanning. Consistently, promoter expression showed no difference between promoters with longer PIC-TSS distance and promoters with shorter PIC-TSS distance in hypothetical *ssl2* processivity GOF alleles.

We assessed if fission yeast *S. pombe* could also use the scanning mechanism for TSS selection through potentially altering Pol II's catalytic activity as a probe for initiation sensitivity. If *S. pombe* also used scanning for TSS selection, we expected to see polar shifts in TSS distributions in *S. pombe* Pol II catalytic mutants as we observed in budding yeast Pol II catalytic mutants. Two Pol II alleles were created, *rpb1*+ E1106G and N1085S, which are homologous to the budding yeast Pol II alleles *rpb1* E1103G and N1082S that harbor increased and decreased catalytic activity, respectively. Assessment of TSS usage in *rpb1*+ E1106G and N1085S at a few model genes and genome-wide revealed a more complex TSS usage pattern in *S. pombe* than what would be predicted under a strict scanning model. *S. pombe rpb1*+ E1106G did not show activation of upstream sites or polar effects on TSS distributions at a few tested model genes. Though TSS distribution at some promoters in the genome showed shift in median TSS position upstream as detected by deep sequencing, this shift was not as significant as our observations in corresponding *S. cerevisiae* Pol II mutants, compared by the examination of the median TSS shift of all promoters. There are two possibilities for this observation, first, *S. pombe* does not use scanning at all promoters for TSS selection; second *rpb1*+ E1106G allele in *S. pombe* doesn't have increased catalytic activity. To determine the latter, experimental measurement of *rpb1*+ E1106G catalytic activity would be required. *S. pombe rpb1*+ N1085S allele altered TSS distributions at some of the tested model genes and genome-wide, with the trend being a downstream shift. However, the patterns of *S. pombe* N1085S shifts were more complex that its *S. cerevisiae* homolog N1082S. *S. pombe* N1085S did not show apparent activation for downstream sites, this difference could result from the different promoter architectures between the two organisms. Alternatively, it could also be due to the putatively decreased Pol II catalytic activity in *S. pombe* N1085S being less than its *S. cerevisiae* homolog N1082S. Further investigation will be needed to examine if Pol II is capable

of scanning for TSS in *S. pombe*. For example, to test if there is a polar TSS shift upon disruption of one of two identical Inrs as has been done in classic scanning experiments in *S. cerevisiae*, or examining the *S. pombe* PIC in single-molecule experiment where downstream DNA movement can be directly observed[20,44]. We used the same rationale as used in *S. pombe* to test if Pol II could also be capable of scanning in the fruit fly *D. melanogaster* **(Appendix A)**. We aimed to create Pol II catalytic activity mutants in *D. melanogaster* that were homologous to *S. cerevisiae* Pol II mutants and tested their effects on TSS usage. Our initial strategy for creating Pol II catalytic activity mutants in *D. melanogaster* was through rescue of a *RpII215* null mutant harboring a mutation on X-chromosome by either WT or mutant *RpII215* transgene introduced on the second chromosome. However, our WT transgene could only complement in males but not in females. To eliminate the location effect for causing the unsuccessfully rescued female, future studies will create Pol II catalytic mutants by site directed genome editing at the native locus through CRISPR.

## 6.2 FUTURE DIRECTIONS

Our first future question will be "is there any competition between Pol II and +1 nucleosome positioning during scanning and, if so, how?". Nucleosomes are composed of DNA wrapped histone proteins and generally serve as transcription repressors[132]. Activation of gene expression requires depletion of nucleosomes, which are usually facilitated by chromatin remodelers. In yeast, RSC is the major remodeler complex that creates a nucleosome-free region (NFR) flanked by -1 and +1 nucleosomes with TSSs residing adjacent to or just within the +1 nucleosome[133]. Studies have shown that depletion of Sth1, the catalytic subunit of RSC, cause NFR encroachment by nucleosomes and changes of TSS utilization[134,135], suggesting a fraction of

TSSs becoming sensitive to nucleosome competition. Ssl2 is located on the downstream edge of the PIC and is presumed to pump downstream DNA toward the upstream Pol II active site. Thus, we predict that there might be some competition between Ssl2 and the +1 nucleosome for DNA accessibility during transcription initiation **(Figure 52)**.

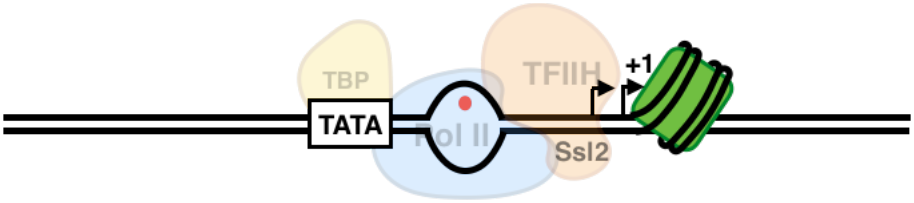



**Figure 52 The hypothetical competition between Ssl2 and +1 nucleosome in DNA accessibility**

Schematic showing of the hypothetical competition between Ssl2 and +1 nucleosome. Ssl2 localizes downstream inside of them PIC and translocates downstream DNA to upstream for Pol II transcription. The +1 nucleosome is downstream of flanking the NFR and has TSS inside of it. The hypothesis is that there is some competition between Ssl2 and the +1 nucleosome for DNA accessibility during transcription initiation.

To test this hypothesis, we performed two simple experiments. First, we did a genetic interaction test between the *ssl2* alleles that affect TSS usage and two histone mutants that contain mutations at the entry/exit sites of the DNA template (histone mutants are from Dr. Karen Arndt's lab, University of Pittsburgh). These two histone alleles are *hht2-T45A* and *hht2-R52A*. We first observed that there were no strong genetic interactions (lethality) as we observe between some of other initiation factors discussed in Chapter III **(Figure 53A)**. Next, we tested the transcription-related phenotypes in these two alleles and found that both of them exhibit moderate level of growth defects but not transcription-related phenotypes **(Figure 53B)**. Compared to the single mutants, double mutants that combine both *ssl2* and histone alleles exhibited no growth difference under tested transcriptional conditions. Therefore, there are no genetic interactions between the

tested histone mutants and *ssl2* mutants. In our future research, we will screen for histone mutants that have altered TSS usage and use them to test for genetic interactions with *ssl2* TSS mutants.

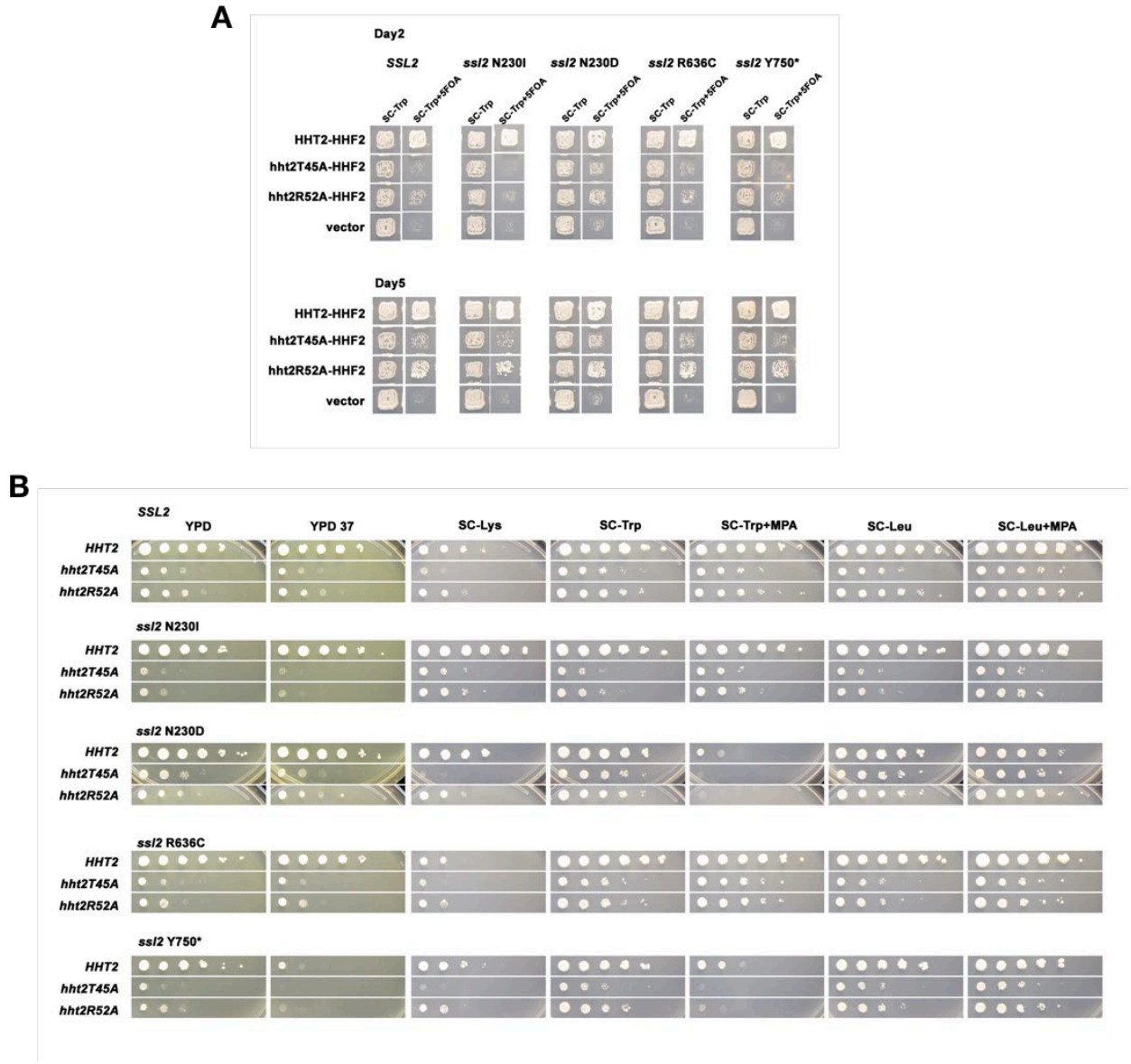



**Figure 53 Genetic interactions between *ssl2* and two histone alleles**

**(A)** No strong genetic interactions (lethality phenotype) are observed between *ssl2* and histone alleles. Patch assay shows that double mutants of histone *hht2T45A/hht2R52A* and *ssl2* N230D, N230I, R636C or Y750* show normal growth phenotypes compared to the correspondent single mutants. **(B)** Patch assay shows that the phenotypes of double mutants and single mutants of histone *hht2T45A/hht2R52A* and *ssl2* N230D, N230I, R636C or Y750* show no difference under tested transcriptional conditions.

Previous study found that depletion of Sth1 caused both +1 nucleosomes and TSS distributions to shift upstream[135]. A very recent study of chromatin remodelers ISW2 and INO80 showed that these factors had effects on promoter nucleosome positioning and transcription start site usage[136]. A ruler model has been proposed that TSS positions may be determined by nucleosome positioning[137]. This model assumes that the width of the TSS window will not be changed in spite of upstream or downstream +1 nucleosome and TSS distribution shift. Alternatively, TSS distributions may be sensitive to nucleosome positioning but not determined by them. Avital Klein-Brill *et. al* tested this model by examining the newly activated TSS upon depletion of Sth1 and found that the number of repressed TSS downstream due to upstream shifted +1 nucleosome is more than the newly activated TSS upstream[135]. However, this observation is not against the hypothesis of a moving TSS window, there might be something else preventing new TSS showing upstream, for example, the downstream movement of -1 nucleosome. We used their data and tested the ruler model through examining the width of the spread of TSS window before and after Sth1 depletion as we analyzed in Chapter IV, and found that the width of TSS window is increased after 30 minutes and 1 hour of Sth1 depletion.

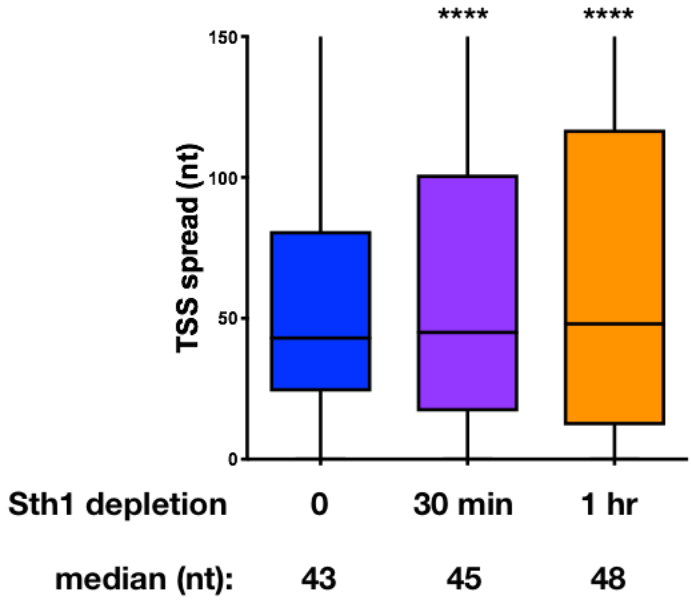



**Figure 54 TSS spread in Sth1 depletion strains**

Boxplot showing the TSS spread in Sth1 depleted 0, 30 minutes and 1 hour strains. There is a significant increase in TSS spread after 30 minutes and 1 hour of Sth1 depletion.

In our future study, we will combine *ssl2* alleles that shift TSS distribution with hypothetically altered processivity and *sth1* depletion allele that shift TSS distribution upstream to see if there are any interactions between the two. **(1)** Our hypothesis is that *ssl2* processivity GOF alleles will show epistasis or suppressive interactions with the *sth1* depletion strain. One model is that *ssl2* processivity GOF allele drives Pol II scanning further downstream by pumping more DNA into the upstream Pol II active site, in contrast, *sth1* depletion allele decreases downstream DNA accessibility by encroachment of the +1 nucleosome on the downstream edge of NFR. The TSS distribution in *ssl2* processivity GOF allele and *sth1* depletion allele combined double mutant will be determined. If the inhibition of TSSs observed in the *sth1* allele are maintained even in the GOF *ssl2* allele it would suggest that putative increase in Ssl2 activity cannot effectively compete

with nucleosomes for TSSs that are nucleosome-sensitive. Alternatively, downstream shift in TSS distribution would be observed if *ssl2* processivity successfully can counteract upstream nucleosome movement of the *sth1* degron strain. Another model is that, in the double mutant, *ssl2* processivity GOF allele's effect on getting more downstream DNA and *sth1* depletion allele's effect on losing downstream DNA compensate to each other and result in WT-like TSS distribution. **(2)** For interactions between the *ssl2* processivity LOF allele and the *sth1* depletion allele, we predict there will be an enhancement of upstream shifts in the double mutants, but we can ask if promoters shown to have *sth1* sensitive TSSs show a greater effect upon combination with *ssl2* LOF versus promoters that do not show *sth1* effects. Such a result would be consistent with putatively reduced Ssl2 processivity mutants being sensitive to promoters with nucleosomes closer to or adjacent to TSSs than promoters without.

# APPENDIX A PRELIMINARY INVESTIGATION OF THE SCANNING MECHANISM IN DROSOPHILA MELANOGASTER

We also tested if scanning is a mechanism used in *Drosophila melanogaster* using the same rationale as used in *S. pombe*. We used two mutants, C4 and S1, containing mutations in the largest and the second largest subunit of *D. melanogaster* Pol II having altered Pol II catalytic activity and tested their TSS usage in 3rd instar larvae by primer extension. Additionally, we created Pol II TL mutants that are analogous to Pol II catalytic mutants in *S. cerevisiae* using a traditional integrase-based transgenic method. However, we found that the transgenic WT gene we introduced cannot fully rescue a null Pol II allele. We therefore switched to CRISPR to directly edit the Pol II TL in *D. melanogaster* genome at its native location.

## A.1 TSS USAGE OF POL II MUTANTS WITH POTENTIALLY ALTERED CATALYTIC ACTIVITY IN *D. MELANOGASTER*

*Drosophila melanogaster* lines used in Appendix A are listed in **Table 17**.

**Table 17 *Drosophila melanogaster* lines used in Appendix A**

| Line number | Genotype |
|---|---|
| 3663 | $v^1$ RpII215$^4$ |
| 34757 | $ras^1$ $v^1$ RpII215$^{K1}$; RpII140$^{S1}$ |
| 11547 | $w^{67c23}$ P{lacW}RpII215$^{G0040}$/FM7c |
| 6328 | $y^1$ $ac^{Hw-1}$ $v^1$ RpII215$^{Ubl}$/FM7c |
| 438 | RpS5a$^2$/FM6; $sna^{Sco}$ Aats-asn$^4$ $pr^1$/CyO |
| DB2 | $v^1$; Sp/CyO; TM2/TM6B |
| 25709 | $y^1$ $v^1$ P{nos-phiC31\int.NLS}X; P{CaryP}attP40 |

To examine if *D. melanogaster* TSSs are similarly sensitive to putative perturbation of Pol II activity in way predicted from the scanning model, we utilized two Pol II mutants in *D. melanogaster* that have literature reported altered *in vitro* elongation rate and tested their TSS usage at selected gene promoters. These two Pol II mutants are C4 and S1. It was reported that C4 had a slower and S1 had a faster than WT *in vitro* elongation rate[138]. The C4 mutant was reported to contain an Arg to His amino acid substitution at position 741 on the largest Pol II subunit RpII215, correspondent to *S. cerevisiae* Rpb1 subunit. The S1 mutant was reported to have a Ser to Cys amino acid change at position 728 in the second largest Pol II subunit RpII140, corresponding to *S. cerevisiae* Rpb2. We first ordered two *Drosophila* lines that contain the reported C4 and S1 alleles from the Flybase (https://flybase.org/) and confirmed their reported genotypes by PCR and sequencing. These two lines are 3663 (Flybase number) that contains a reported C4 mutation and 34757 that has a reported S1 mutation and an additional K1 mutation, as listed in **Table 18**. Our genotyping results showed that 3663 indeed contained a C4 mutation (R741H), however, 34757 contained reported K1 mutation but not S1 mutation. Instead, an S11 mutation (M735V) was detected in 34757. Because of the missing S1 mutation, we continued our next analysis only on 3663 that contains the C4 mutation.

**Table 18 Genotyping of Pol II mutants used in the literature**

| Line number | Allele | Reported mutation | Sequenced mutation |
|---|---|---|---|
| **3663** | C4 | RpII215_C4, 741: ARG to His | RpII215_C4, 741: ARG to His |
| **34757** | K1 | RpII215_K1, 678: Ser to Asn | RpII215_K1, 678: Ser to Asn |
| | S1 | RpII140_S1, 728: Ser to Cys | RpII140_S11, 735: Met to Val |

We next examined the C4 allele's effect on TSS usage by primer extension at *RPL19* promoter using 3rd instar larvae extracted RNA **(Figure 55A, B)**. We observed that C4 allele with reported slower elongation rate affected TSS usage at *RPL19* promoter, but not in a way that is similar to Pol II LOF allele affect TSS usage in *S. cerevisiae*. Primer extension result shows that the overall TSS usage is increased in C4 mutant, but there is no clear pattern of TSS polar shift.

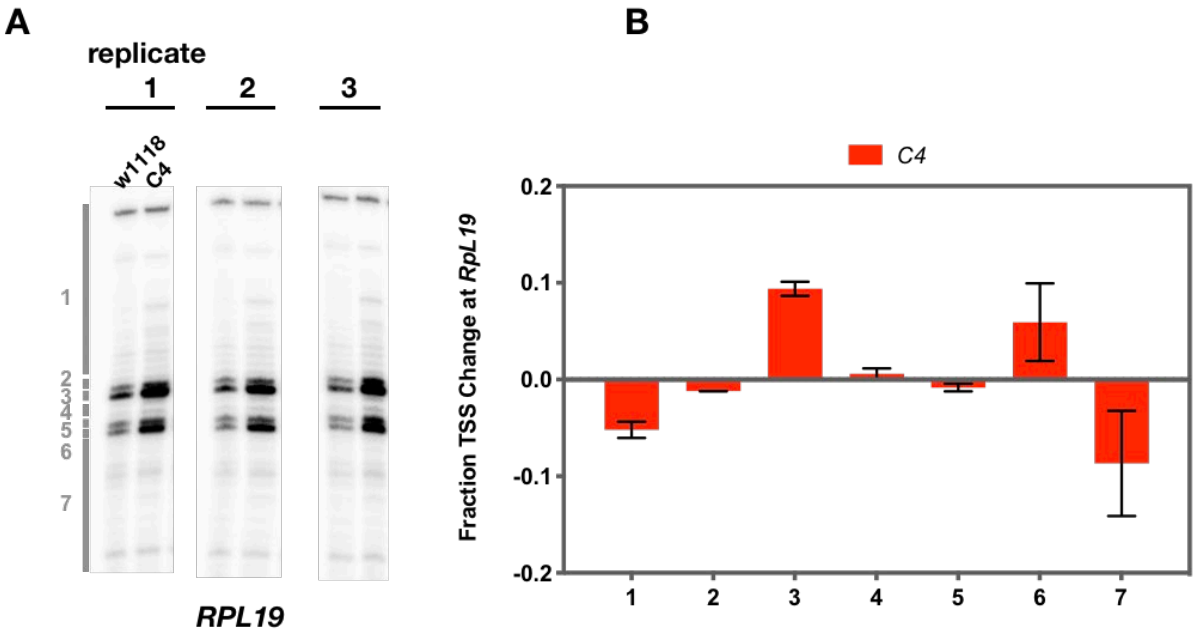



**Figure 55 TSS usage of *D. melanogaster* C4 mutant at *RPL19* promoter**

**(A)** *D. melanogaster* C4 mutant affect TSS usage at *RPL19* promoter. Primer extension result showed that the overall TSS is increased in C4 mutants, but no clear shift is visually observed. **(B)** Quantification of TSS usage at *RPL19* promoter promoter in *D. melanogaster* C4 mutant. TSS usage in bin 3 and 6 is increased, whearas TSS usage in bin 1 and 7 is decreased.

## A.2 TRANSGENIC *RPII215* GENE WE INTRODUCED CANNOT FULLY RESCUE A NULL POL II ALLELE

To create Pol II TL mutants that are analogous to Pol II catalytic mutants in *S. cerevisiae*, we first used a traditional integrase-based transgenic method. Our strategy was to introduce the WT/mutant copy of *RpII215*$^{+}$ on the second chromosome of a *D. melanogaster* strain that harbors a *RpII215* null mutation on the X chromosome **(Figure 56)**. The detailed strategy and *D. melanogaster* cross steps are illustrated in **Figure 57**. We used two *RpII215* null alleles 11547 and 6328, one contains a P-element and the other contains a point mutation on *RpII215* that both disrupt *RpII215*'s function. We found that one copy of transgenic WT *RpII215*$^{+}$ gene we introduced on the second chromosome plus one original copy of WT *RpII215*$^{+}$ on the second chromosome are viable (**Figure 57, offspring of cross 6**). Homozygous transgenic strain that have two copies of introduced WT *RpII215*$^{+}$ is male viable but female inviable, with females dying at the pupal stage. We are not sure if the transgenic genes' location affected the rescue. We are currently collaborating with a research group to use CRISPR to edit the wanted TL site on *D. melanogaster* genome at its native location.

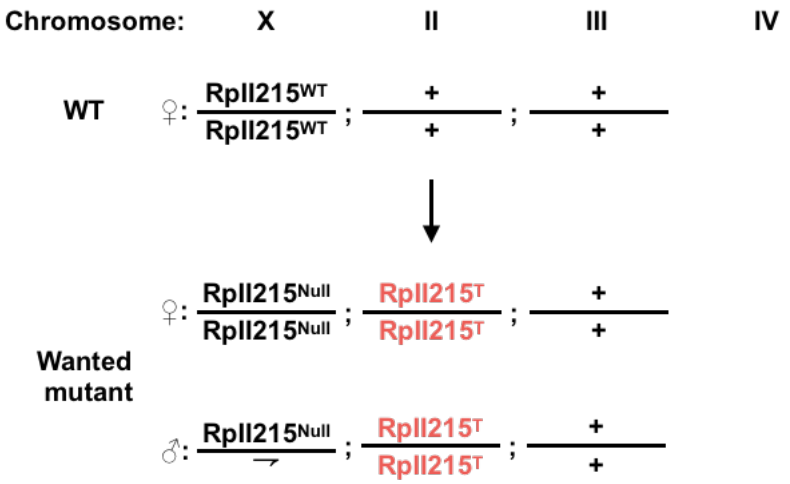
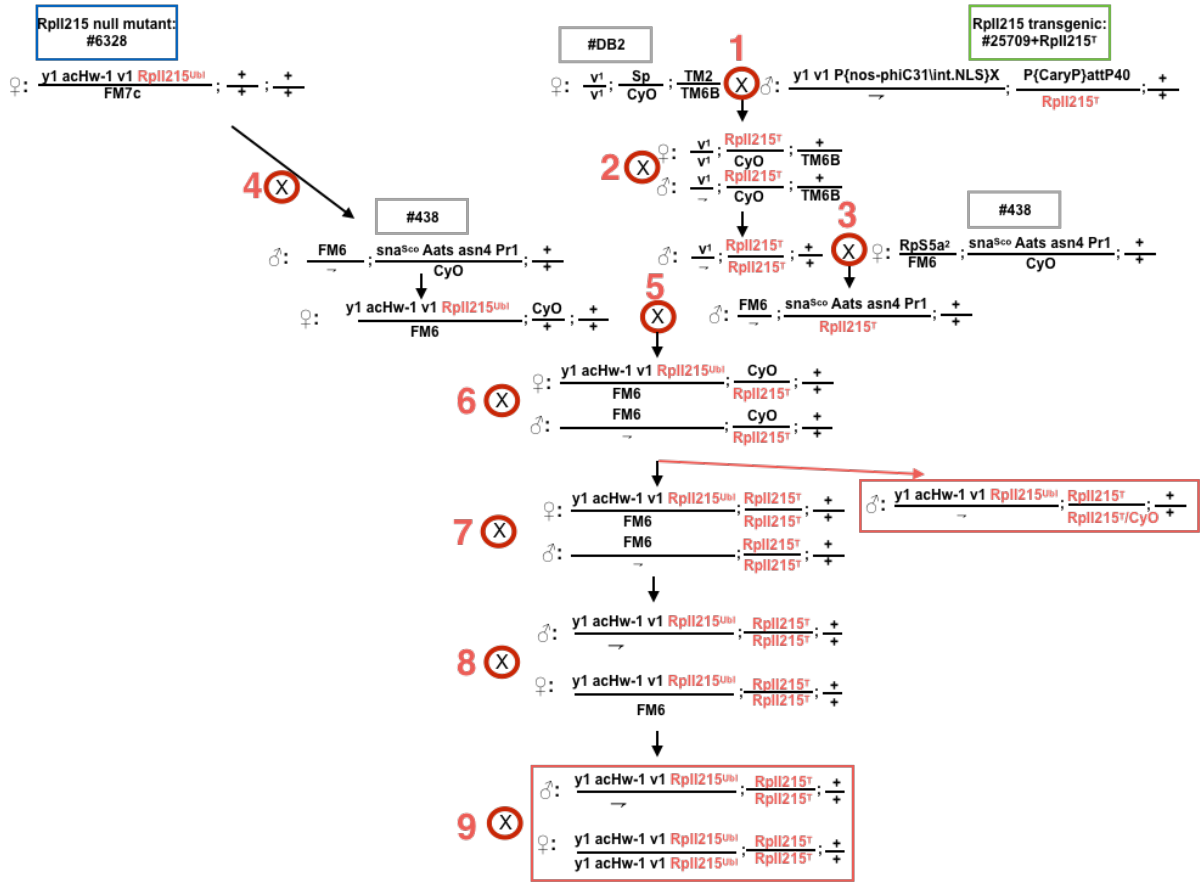



**Figure 56 Overview of *D. melanogaster* transgenic strategy**

Our strategy was to introduce a WT/mutant copy of *RpII215* on the second chromosome of a *RpII215* null mutant, through fly crossing, obtain a rescued line that has *RpII215* null allele on X chromosome and the introduced copies, either WT or mutant copy, on the second chromosome.

**A**


RpII215 null mutant: #6328
RpII215 transgenic: #25709+RpII215T
#DB2
#438
#438
1
2
3
4
5
6
7
8
9

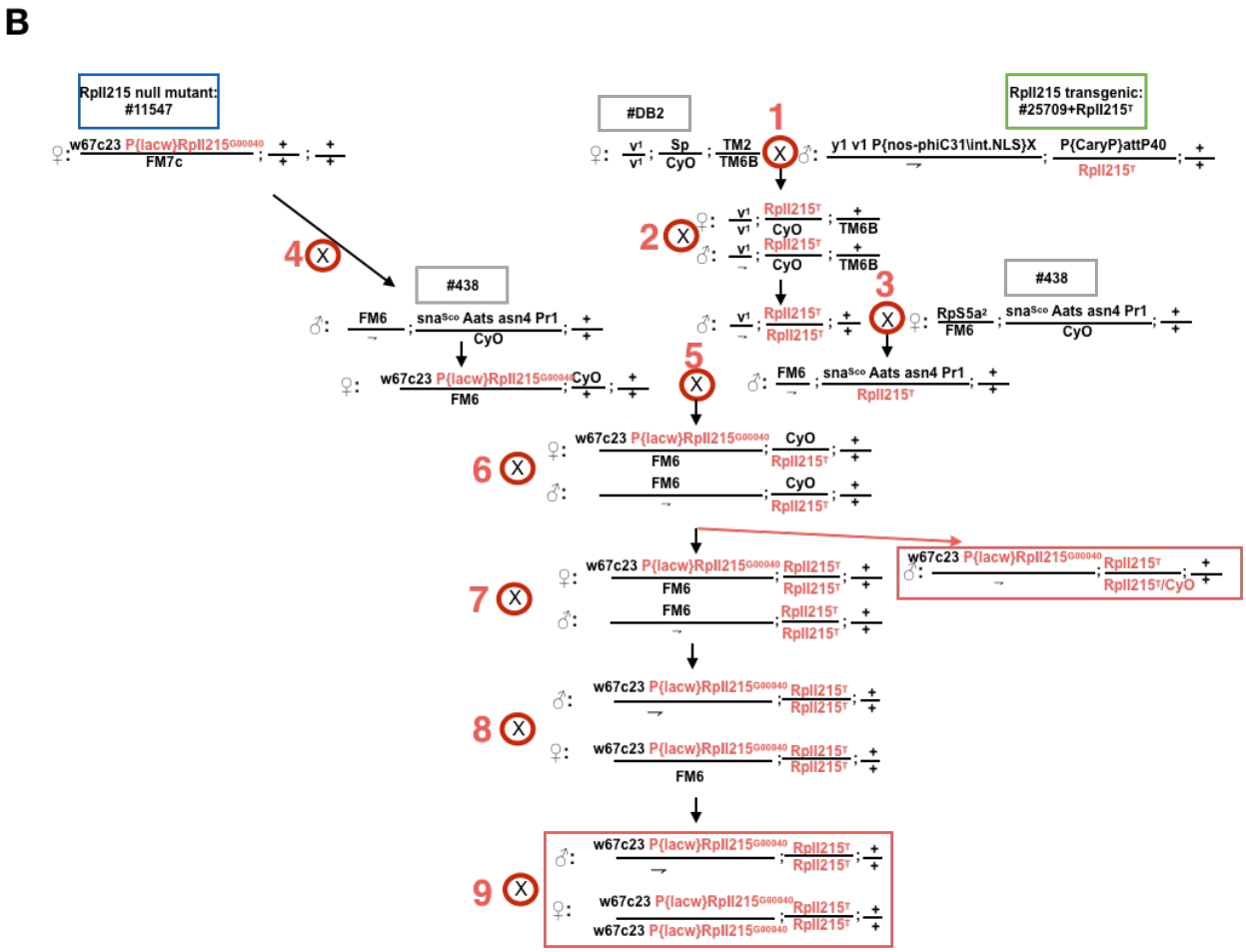



**Figure 57 Fly crossing strategy for obtaining homozygous transgenic lines**

We introduce an *in vitro* cloned WT *RpII215*$^{+}$ in to the second chromosome of fly line 25709 through integrase-based injection method, the injection was performed by Rainbow Transgenic Flies (https://www.rainbowgene.com/). Offspring of the trangenic line that has succeful trangenic gene integrated was selected based on the eye color marker and crossed with DB2, followed by a serial of crossing steps as indicated above to obtain wanted rescue lines. As stated in main chapter, mutant line that contains one copy of transgenic WT *RpII215*$^{+}$ gene plus an original copy of WT *RpII215*$^{+}$ on the second chromosome are viable (offspring of cross 6, one in red box). However, homozygous transgenic strain that have two copies of introduced WT *RpII215*$^{+}$ is male viable but female inviable, female dies at pupal stage (offspring of cross 8).

# BIBLIOGRAPHY


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

2. Kostrewa, D. et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323-30 (2009).

3. Grunberg, S. & Hahn, S. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* **38**, 603-11 (2013).

4. Tirode, F., Busso, D., Coin, F. & Egly, J.-M. Reconstitution of the Transcription Factor TFIIH: Assignment of Functions for the Three Enzymatic Subunits, XPB, XPD, and cdk7. *Molecular Cell* **3**, 87-95 (1999).

5. Fishburn, J., Tomko, E., Galburt, E. & Hahn, S. Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proc Natl Acad Sci U S A* **112**, 3961-6 (2015).

6. Kim, T.-K., Ebright, R.H. & Reinberg, D. Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH. *Science* **288**, 1418-1421 (2000).

7. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204-209 (2017).

8. Butler, J.E.F. & Kadonaga, J.T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & development* **16**, 2583-2592 (2002).

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

10. Yang, C., Bolotin, E., Jiang, T., Sladek, F.M. & Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**, 52-65 (2007).

11. Basehoar, A.D., Zanton, S.J. & Pugh, B.F. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell* **116**, 699-709 (2004).

12. Reeve, J.N. Archaeal chromatin and transcription. *Molecular Microbiology* **48**, 587-598 (2003).

13. Yamagishi, J. et al. High-resolution characterization of toxoplasma gondii transcriptome with a massive parallel sequencing method. *DNA Research* **17**, 233-243 (2010).

14. Donczew, R. & Hahn, S. Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters. *Mol Cell Biol* **38**(2018).

15. Purnell, B.A., Emanuel, P.A. & Gilmour, D.S. TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. *Genes & development* **8**, 830-842 (1994).

16. Chalkley, G.E. & Verrijzer, C.P. DNA binding site selection by RNA polymerase II TAFs: a TAFII250–TAFII150 complex recognizes the Initiator. *The EMBO Journal* **18**, 4835-4845 (1999).

17. Carninci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**, 626-35 (2006).

18. Frith, M.C. et al. A code for transcription initiation in mammalian genomes. *Genome Research* **18**, 1-12 (2008).

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

20. Kuehner, J.N. & Brow, D.A. Quantitative Analysis of in Vivo Initiator Selection by Yeast RNA Polymerase II Supports a Scanning Model. *Journal of Biological Chemistry* **281**, 14119-14128 (2006).

21. Basu, R.S. et al. Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *The Journal of biological chemistry* **289**, 24549-24559 (2014).

22. Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D. & Ebright, R.H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & development* **12**, 34-44 (1998).

23. Evans, R., Fairley, J.A. & Roberts, S.G. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes & development* **15**, 2945-2949 (2001).

24. Deng, W. & Roberts, S.G.E. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & development* **19**, 2418-2423 (2005).

25. Burke, T.W. & Kadonaga, J.T. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes & development* **10**, 711-724 (1996).

26. Burke, T.W. & Kadonaga, J.T. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & development* **11**, 3020-3031 (1997).

27. Reynolds, G.A. et al. HMG CoA reductase: A negatively regulated gene with unusual promoter and 5′ untranslated regions. *Cell* **38**, 275-285 (1984).

28. Hoskins, R.A. et al. Genome-wide analysis of promoter architecture in Drosophila melanogaster. *Genome Res* **21**, 182-92 (2011).

29. Struhl, K. Molecular Mechanisms Of Transcriptional Regulation In Yeast. *Annual Review of Biochemistry* **58**, 1051-1077 (1989).

30. Hume, D.A. et al. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Reviews Genetics* **8**, 424-436 (2007).

31. Müller, F. & Tora, L. Chromatin and DNA sequences in defining promoters for transcription initiation. *BBA - Gene Regulatory Mechanisms* **1839**, 118-128 (2014).

32. Thomas, M.C. & Chiang, C.M. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**, 105-78 (2006).

33. Sikorski, T.W. & Buratowski, S. The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* **21**, 344-51 (2009).

34. Hahn, S. & Young, E.T. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189**, 705-36 (2011).

35. Anandapadamanaban, M. et al. High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nature structural & molecular biology* **20**, 1008-1014 (2013).

36. Knoll, E.R. ProQuest Dissertations Publishing (2017).

37. Huisinga, K.L. & Pugh, B.F. A Genome-Wide Housekeeping Role for TFIID and a Highly Regulated Stress-Related Role for SAGA in Saccharomyces cerevisiae. *Molecular Cell* **13**, 573-585 (2004).

38. Baptista, T. et al. SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Molecular Cell* **68**, 130-143.e5 (2017).

39. Warfield, L. et al. Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. *Molecular Cell* **68**, 118-129.e5 (2017).

40. Wang, W., Carey, M. & Gralla, J.D. Polymerase II Promoter Activation: Closed Complex Formation and ATP-Driven Start Site Opening. *Science* **255**, 450-453 (1992).

41. Dvir, A. et al. A Role for ATP and TFIIH in Activation of the RNA Polymerase II Preinitiation Complex Prior to Transcription Initiation. *Journal of Biological Chemistry* **271**, 7245-7248 (1996).

42. Holstege, F.C.P. & Timmers, H.T.M. Analysis of Open Complex Formation during RNA Polymerase II Transcription Initiation Using Heteroduplex Templates and Potassium Permanganate Probing. *Methods* **12**, 203-211 (1997).

43. Giardina, C. & Lis, J.T. DNA melting on yeast RNA polymerase II promoters. *Science (New York, N.Y.)* **261**, 759-762 (1993).

44. Fazal, F.M., Meng, C.A., Murakami, K., Kornberg, R.D. & Block, S.M. Real-time observation of the initiation of RNA polymerase II transcription. *Nature* **525**, 274-7 (2015).

45. Luo, J. et al. Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. *Mol Cell* **59**, 794-806 (2015).

46. Jesper Q. Svejstrup, W.J.F., Janice LaPointe, and Roger D. Kornberg. RNA Polymerase Transcription FactorIIH Holoenzyme from Yeast. *The Journal of Biological Chemistry* **289**, 28044-28048 (1994).

47. Takagi, Y. et al. Revised subunit structure of yeast transcription factor IIH (TFIIH) and reconciliation with human TFIIH. *J Biol Chem* **278**, 43897-900 (2003).

48. Grunberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol* **19**, 788-96 (2012).

49. Murakami, K. et al. 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**, 4816-21 (2012).

50. Jesper Q Svejstrup, Z.W., William J Feave, Xiahua Wu, David A Bushnell, Thomas F Donahue, Errol C Friedberg, Roger D Kornberg. Different forms of TFIIH for transcription and DNA repair: Holo-TFIIH and a nucleotide excision repairosome. *Cell* **80**, 21-28 (1995).

51. Compe, E. & Egly, J.M. TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol* **13**, 343-54 (2012).

52. Greber, B.J. et al. The complete structure of the human TFIIH core complex. *eLife* **8**(2019).

53. Gulyas, K.D. & Donahue, T.F. SSL2, a suppressor of a stem-loop mutation in the HIS4 leader encodes the yeast homolog of human ERCC-3. *Cell* **69**, 1031-1042 (1992).

54. Qiu, H., Park, E., Prakash, L. & Prakash, S. The Saccharomyces cerevisiae DNA repair gene RAD25 is required for transcription by RNA polymerase II. *Genes and Development* **7**, 2161-2171 (1993).

55. Egly, J.-M., Bergmann, E., Tremeau-Bravard, A. & Coin, F. Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. *The EMBO Journal* **18**, 1357-1366 (1999).

56. Goel, S., Krishnamurthy, S. & Hampsey, M. Mechanism of start site selection by RNA polymerase II: interplay between TFIIB and Ssl2/XPB helicase subunit of TFIIH. *J Biol Chem* **287**, 557-67 (2012).

57. Hahn, S., Hoar, E.T. & Guarente, L. Each of Three "TATA Elements" Specifies a Subset of the Transcription Initiation Sites at the CYC-1 Promoter of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 8562-8566 (1985).

58. Singleton, M.R., Dillingham, M.S. & Wigley, D.B. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**, 23-50 (2007).

59. Tuteja, N. & Tuteja, R. Unraveling DNA helicases. Motif, structure, mechanism and function. *Eur J Biochem* **271**, 1849-63 (2004).

60. Lohman, T.M., Tomko, E.J. & Wu, C.G. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol* **9**, 391-401 (2008).

61. Fairman-Williams, M.E., Guenther, U.P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* **20**, 313-24 (2010).

62. Yang, Q., Del Campo, M., Lambowitz, A.M. & Jankowsky, E. DEAD-box proteins unwind duplexes by local strand separation. *Mol Cell* **28**, 253-63 (2007).

63. Myong, S. et al. Cytosolic Viral Sensor RIG-I Is a 5′ -Triphosphate-Dependent Translocase on Double-Stranded RNA. *Science* **323**, 1070-1074 (2009).

64. Fan, L. et al. Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol Cell* **22**, 27-37 (2006).

65. Hilario, E., Li, Y., Nobumori, Y., Liu, X. & Fan, L. Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. *Acta Crystallogr D Biol Crystallogr* **69**, 237-46 (2013).

66. Greber, B.J. et al. The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549**, 414-417 (2017).

67. Kokic, G. et al. Structural basis of TFIIH activation for nucleotide excision repair. *Nat Commun* **10**, 2885 (2019).

68. Yan, C. et al. Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. *Nat Struct Mol Biol* **26**, 397-406 (2019).

69. Egly, J.-M., de Jesus, B.B., Zhovmer, A., Oksenych, V. & Coin, F. Molecular insights into the recruitment of TFIIH to sites of DNA damage. *The EMBO Journal* **28**, 2971-2980 (2009).

70. Jawhari, A. et al. p52 Mediates XPB function within the transcription/repair factor TFIIH. *J Biol Chem* **277**, 31761-7 (2002).

71. Cleaver, J.E., Thompson, L.H., Richardson, A.S. & States, J.C. A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Human mutation* **14**, 9-22 (1999).

72. Murakami, K. et al. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724 (2013).

73. Robinson, P.J. et al. Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**, 1411-1422 e16 (2016).

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

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

76. Cramer, P. et al. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37**, 337-52 (2008).

77. Braberg, H. et al. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* **154**, 775-88 (2013).

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

79. Qiu, C. et al. Promoter scanning during transcription initiation in Saccharomyces cerevisiae: Pol II in the "shooting gallery". *bioRxiv*, 810127 (2019).

80. Chen, B.S. & Hampsey, M. Functional interaction between TFIIB and the Rpb2 subunit of RNA polymerase II: implications for the mechanism of transcription initiation. *Mol Cell Biol* **24**, 3983-91 (2004).

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

82. Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437-40 (2013).

83. Chen, H. et al. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119-128 (1995).

84. Tsai, F.T.F. & Sigler, P.B. Structural basis of preinitiation complex assembly on human Pol II promoters. *The EMBO Journal* **19**, 25-36 (2000).

85. Thomas, M.C. & Chiang, C.-M. The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology* **41**, 105-178 (2006).

86. Faitar, S.L., Brodie, S.A. & Ponticelli, A.S. 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 (2001).

87. Čabart, P., Újvári, A., Pal, M. & Luse, D.S. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15786-15791 (2011).

88. Kilpatrick, A.M., Koharudin, L.M.I., Calero, G.A. & Gronenborn, A.M. Structural and binding studies of the C‐terminal domains of yeast TFIIF subunits Tfg1 and Tfg2. *Proteins: Structure, Function, and Bioinformatics* **80**, 519-529 (2012).

89. Plaschka, C. et al. Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353-358 (2016).

90. Sun, Z.-W. & Hampsey, M. Identification of the Gene (SSU71/TFG1) Encoding the Largest Subunit of Transcription Factor TFIIF as a Suppressor of a TFIIB Mutation in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3127-3131 (1995).

91. Freire-Picos, M.A., Krishnamurthy, S., Sun, Z.W. & Hampsey, M. 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**, 5045-52 (2005).

92. Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M. & Ponticelli, A.S. Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Mol Cell Biol* **24**, 10975-85 (2004).

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

94. Kretzschmar, M., Kaiser, K., Lottspeich, F. & Meisterernst, M. A novel mediator of class II gene transcription with homology to viral immediate-early transcriptional regulators. *Cell* **78**, 525-534 (1994).

95. Henry, N.L., Bushnell, D.A. & Kornberg, R.D. A Yeast Transcriptional Stimulatory Protein Similar to Human PC4. *Journal of Biological Chemistry* **271**, 21842-21847 (1996).

96. Sikorski, Timothy W. et al. Sub1 and RPA Associate with RNA Polymerase II at Different Stages of Transcription. *Molecular Cell* **44**, 397-409 (2011).

97. James E. Cleaver, L.H.T., 2 Audrey S. Richardson,3 and J. Christopher States3. A Summary of Mutations in the UV-Sensitive Disorders: Xeroderma Pigmentosum, Cockayne Syndrome, and Trichothiodystrophy. *HUMAN MUTATION* **14**, 9-22 (1999).

98. Oh, K.S. et al. Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Hum Mutat* **27**, 1092-103 (2006).

99. Weeda, G. et al. A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell* **62**, 777-791 (1990).

100. Sweder, K.S. & Hanawalt, P.C. The COOH terminus of suppressor of stem loop (SSL2/RAD25) in yeast is essential for overall genomic excision repair and transcription-coupled repair. *Journal of Biological Chemistry* **269**, 1852-1857 (1994).

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

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

103. Amberg DC, B.D., Strathern JN. Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition. . *Cold Spring Harbor, NY: Cold Spring Harbor Press.* , 230 p. (2005).

104. Ranish, J.A. & Hahn, S. The yeast general transcription factor TFIIA is composed of two polypeptide subunits. *J Biol Chem* **266**, 19320-7 (1991).

105. Kuehner, J.N. & Brow, D.A. Regulation of a Eukaryotic Gene by GTP-Dependent Start Site Selection and Transcription Attenuation. *Molecular Cell* **74**, 634-634 (2019).

106. Simchen, G., Winston, F., Styles, C.A. & Fink, G.R. Ty-Mediated Gene Expression of the LYS2 and HIS4 Genes of Saccharomyces cerevisiae is Controlled by the Same SPT Genes. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 2431-2434 (1984).

107. Simchen, G., Winston, F., Styles, C.A. & Fink, G.R. Ty-mediated gene expression of the LYS2 and HIS4 genes of Saccharomyces cerevisiae is controlled by the same SPT genes. *Proceedings of the National Academy of Sciences* **81**, 2431-2434 (1984).

108. He, Y., Fang, J., Taatjes, D.J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481-6 (2013).

109. Boeger, H., Griesenbeck, J., Strattan, J.S. & Kornberg, R.D. Nucleosomes Unfold Completely at a Transcriptionally Active Promoter. *Molecular Cell* **11**, 1587-1598 (2003).

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

111. Wu, W.-H., Pinto, I., Chen, B.-S. & Hampsey, M. Mutational Analysis of Yeast TFIIB: A Functional Relationship Between Ssu72 and Sub1/Tsp1 Defined by Allele-Specific Interactions With TFIIB. *Genetics* **153**, 643-652 (1999).

112. Schmitt, M.E., Brown, T.A. & Trumpower, B.L. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Res* **18**, 3091-2 (1990).

113. Vvedenskaya, I.O., Goldman, S.R. & Nickels, B.E. Preparation of cDNA Libraries for High-Throughput RNA Sequencing Analysis of RNA 5′ Ends. in *Bacterial Transcriptional Control: Methods and Protocols* (eds. Artsimovitch, I. & Santangelo, T.J.) 211-228 (Springer New York, New York, NY, 2015).

114. Vvedenskaya, I.O., Goldman, S.R. & Nickels, B.E. Preparation of cDNA libraries for high-throughput RNA sequencing analysis of RNA 5' ends. *Methods Mol Biol* **1276**, 211-28 (2015).

115. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).

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

117. Subgroup, G.P.D.P. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

118. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

119. Schwalb, B. et al. LSD: Lots of Superior Depictions. (2015).

120. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

121. Lamia Wahba, 3,4 Lorenzo Costantino,1,4 Frederick J. Tan,2 Anjali Zimmer,1 and Douglas Koshland1. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *GENES & DEVELOPMENT* **30:1327–1338**(2016).

122. Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645-3647 (2017).

123. Freese, N.H., Norris, D.C. & Loraine, A.E. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**, 2089-2095 (2016).

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

125. Choi, W.S., Yan, M., Nusinow, D. & Gralla, J.D. In Vitro Transcription and Start Site Selection in Schizosaccharomyces pombe. *Journal of Molecular Biology* **319**, 1005-1013 (2002).

126. Fong, N. et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes & development* **28**, 2663-2676 (2014).

127. Forsburg, S.L. Introduction of DNA into S. pombe cells. *Curr Protoc Mol Biol* **Chapter 13**, Unit 13.17 (2003).

128. Gao, J. et al. Rapid, efficient and precise allele replacement in the fission yeast Schizosaccharomyces pombe. *Current Genetics* **60**, 109-119 (2014).

129. Li, H. et al. Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE. *RNA Biol* **12**, 525-37 (2015).

130. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, W187-W191 (2014).

131. Hoffman, C.S. & Winston, F. A transcriptionally regulated expression vector for the fission yeast Schizosaccharomyces pombe. *Gene* **84**, 473-479 (1989).

132. Lorch, Y. & Kornberg, R.D. Chromatin-remodeling for transcription. *Quarterly reviews of biophysics* **50**, e5 (2017).

133. Raisner, R.M. et al. Histone Variant H2A.Z Marks the 5′ Ends of Both Active and Inactive Genes in Euchromatin. *Cell* **123**, 233-248 (2005).

134. Hartley, P.D. & Madhani, H.D. Mechanisms that Specify Promoter Nucleosome Location and Identity. *Cell* **137**, 445-458 (2009).

135. Klein-Brill, A., Joseph-Strauss, D., Appleboim, A. & Friedman, N. Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. *Cell Reports* **26**, 279-292.e5 (2019).

136. Kubik, S. et al. Opposing chromatin remodelers control transcription initiation frequency and start site selection. *NATURE STRUCTURAL & MOLECULAR BIOLOGY* **26**, 744-744 (2019).

137. Hughes, Amanda L., Jin, Y., Rando, Oliver J. & Struhl, K. A Functional Evolutionary Approach to Identify Determinants of Nucleosome Positioning: A Unifying Model for Establishing the Genome-wide Pattern. *Molecular Cell* **48**, 5-15 (2012).

138. Chen, Y., Chafin, D., Price, D.H. & Greenleaf, A.L. Drosophila RNA Polymerase II Mutants That Affect Transcription Elongation. *Journal of Biological Chemistry* **271**, 5993 (1996).