January 2016

The Role of the Set1 RNA Polymerase II Interacting Motif (SRIM) in Set1 Recruitment and Histone H3K4 Methylation

Yueping Zhang
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yueping Zhang

Entitled

The Role of the Set1 RNA Polymerase II Interacting Motif (SRIM) in Set1 Recruitment and Histone H3K4 Methylation

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Scott D. Briggs Chair	12/5/2016
Elizabeth J. Tran	12/5/2016
Mark C. Hall	12/5/2016
Tony R. Hazbun	11/23/2016

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Scott D. Briggs

Approved by: Andrew D. Mesecar 12/6/2016

Head of the Departmental Graduate Program Date

THE ROLE OF THE SET1 RNA POLYMERASE II INTERACTING MOTIF (SRIM) IN SET1 RECRUITMENT AND HISTONE H3K4 METHYLATION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yueping Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2016

Purdue University

West Lafayette, Indiana

This dissertation is dedicated to my parents.

ACKNOWLEDGEMENTS

I would like to first thank my advisor, Dr. Scott D. Briggs. Without Scott, I could not accomplish this thesis and study at Purdue University. His restrict scientific training helps me. His marathon running inspired me to finish my first half marathon at Purdue. I would like to thank my committee members, Dr. Elizabeth J. Tran, Dr. Mark C. Hall, and Dr. Tony R. Hazbun for their advice during my time here at Purdue. I also should give thanks to all the members of the Briggs Lab, both past and present. That includes Dr. Hai-Ning Du, Dr. Paul F. South, Dr. Kayla Harmeyer, Luyen Nguyen, and Nina Serratore. I must give thanks for the all the members of department of biochemistry for their help during these years. Finally, I would thank my parents for their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Appendix Figure

LIST OF ABBREVIATIONS

5-FOA	5-fluoroorotic acid
6-AU	6-azauracil
Acetyl-CoA	Acetyl coenzyme A
ACT1	ACTin
ATP	Adenosine triphosphate
BFA	Brefeldin A
BRE1	Brefeldin A sensitivity 1
BRE2	Brefeldin A sensitivity 2
CHD1	Chromo-helicase/ATPase-DNA-binding domain protein 1
ChIP	Chromatin Immunoprecipitation
COMPASS	Complex associated with Set1
CTD	Carboxy-terminal domain of RNA polymerase II
DNA	Deoxyribonucleic acid
DOT1	Disruptor of telomeric silencing 1
FACT	Facilitates chromatin transcription
FAD	Flavin adenine dinucleotide
G6PDH	Glucose-6-phosphate dehydrogenase
GAL1	GALactose metabolism protein 1

GST	Glutathione S-transferase
H2Aub	Histone H2A ubiquitination
H2Bub	Histone H2B ubiquitination
H3K27	Histone H3 Lysine 27
H3K36	Histone H3 lysine 36
H3K4	Histone H3 Lysine 4
H3K79	Histone H3 lysine 79
H3K9	Histone H3 lysine 9
H3T3	Histone H3 threonine 3
H4K20	Histone H4 lysine 20
HAT	Histone acetyltransferase
HDAC	Histone deacetlyase
HMT	Histone methyltransferase
HOX	Homeobox
INO80	INOsitol requiring
IP	Immunoprecipitation
kD	Kilo Dalton
KMT	Histone lysine methyltransferase
LSD1	Lysine-specific histone demethylase 1
MDH3	Malate dehydrogenase 3
me1	monomethylation
me2	dimethylation

me3	trimethylation
MLL	Mixed lineage leukemia
MPA	mycophenolic acid
mRNA	Messenger RNA
NuRD	Nucleosome remodeling deacetylase
NURF	Nucleosome remodeling factor
OD600	Optical density at 600nm
PADI4	Peridylarginine deiminase 4
Paf1C	Paf1 complex
PBS(-K)	Phosphate buffered saline without potassium
PCK1	Phosphoenolpyruvate CarboxyKinase 1
PCR	Polymerase chain reaction
PHD	Plant homeo domain
PMA1	Plasma Membrane ATPase 1
PMSF	Phenolmethylsulfonyl fluoride
PRMT	Protein arginine methyltransferase
PTM	Post-translational modification
PYK1	Pyruvate kinase 1
qRT-PCR	Quantitative real time polymerase chain reaction
RAD6	Radiation sensitive 6
rDNA	Ribosomal DNA
RNAPII	RNA polymerase II

RPB3	RNA polymerase II third largest subunit
RRM	RNA recognition motif
SC-ura	Synthetic complete media dropout uracil
SDI	Sdc1 Dpy-30 interacting
SDS PAGE	Sodium Dodecylsulfate polyacrylamide gel electrophoresis
SET	Su(var), Enhancer of zeste, and Trithorax
SET1	SET domain protein 1
Set1C	Set1 complex
SET2	SET domain protein 2
SIR	Silent Information Regulator
SPP1	Set1C, PHD finger protein 1
SPRY	Spore lysis Ryanodine receptor
SRI domain	Set2 Rpb1 interacting domain
SWD1	Set1C, WD40 repeat protein 1
SWD2	Set1C, WD40 repeat protein 2
SWD3	Set1C, WD40 repeat protein 3
SWI/SNF	Switch/Sucrose Non-fermentable
TSS	Transcriptional start site
URA3	Uracil synthesis-requiring gene 3
WT	Wild type
YPD	Yeast extract/peptone/dextrose media

ABSTRACT

Zhang, Yueping. Ph.D., Purdue University, December 2016. The Role of the Set1 RNA Polymerase II Interacting Motif (SRIM) in Set1 Recruitment and Histone H3K4 Methylation. Major Professor: Scott D. Briggs.

In eukaryotes, gene expression is regulated by epigenetic modifications. Among these modifications, histone H3 lysine 4 (H3K4) methylation has been associated with transcription activation. Set1 is the sole H3K4 histone methyltransferase in *Saccharomyces cerevisiae* and functions in a multi-subunit protein complex (Set1C) to catalyze H3K4 mono-, di-, and trimethylation. In addition, mis-regulation of Set1 human orthologs such as *MLL* (1 to 4) or *SETD1A/B* have been linked to a variety of cancers such as leukemia and early embryonic development defects. The current model indicates that the Paf1 complex (RNA polymerase II associated factor I) recruits Set1C to the phosphorylated CTD of RNA polymerase II (RNAPII) and chromatin, so that Set1C can deposit histone H3K4 methylation at actively transcribed genes. Here we show that the Set1 and RNAPII association is independent of Paf1 complex subunits. Our results also show that Set1 and RNAPII can interact independently of Set1-associated subunits. We have identified, for the first time, the region responsible for the Set1 and RNAPII interaction and have named this region the Set1 RNA polymerase II interacting motif (SRIM), which is in the uncharacterized region upstream of the previously identified Set1 RNA Recognition Motifs (RRM). Importantly, we have determined that Set1 interacts

with the Ser5 phosphorylated form of CTD and that the SRIM is necessary for this interaction. Additionally, SRIM mutants show altered enrichment at mostly the 5' regions of open reading frames of actively transcribed genes, including *PYK1* and *PMA1* and inducible genes, including *PCK1*. Moreover, one SRIM mutant shows a great decrease in H3K4 trimethylation levels globally and gene specifically. Interestingly, the deposition of global H3K4 mono- and dimethylation still occurs in the SRIM mutants. Together, these results suggest that the Set1 and RNAPII interaction is important for proper recruitment of Set1 and the deposition of H3K4 trimethylation to 5' ends of actively transcribed genes, whereas Set1 could also be recruited to chromatin through other mechanisms independent of the Set1-RNAPII interaction to deposit H3K4 methylation. Furthermore, the loss of Set1 and SRIM mutants show sensitivity to 6-azauracil (6AU) indicating that the Set1 and RNAPII interaction is important for transcriptional elongation. Taken together, these data indicate that Set1 specifically interacts with the Ser5 phosphorylated form of RNAPII through the newly identified SRIM and that this interaction plays a role in the establishment of Set1-mediated H3K4 methylation to actively transcribing genes and inducible genes.

In the meanwhile, we have constructed the "N-ICE" plasmids so that non-essential and essential genes can be N-terminally 3×FLAG tagged and expressed from an Inducible promoter (*GAL1*), Constitutive promoters (*CYC1* or *PYK1*), or the Endogenous promoter. We have validated the N-ICE plasmid system by N-terminal tagging two non-essential genes (*SET1* and *SET2)* and two essential genes (*ERG11* and *PKC1*).

.

CHAPTER 1. INTRODUCTION

1.1 Epigenetics and Chromatin

1.1.1 Definition of Epigenetics

The term “epigenetics” was first proposed by Dr. Conrad Hal Waddington when he studied embryogenesis in 1942 to define the study of causal mechanisms connecting genotype to phenotype (1, 2). In recent years, epigenetics has become an area of rapid growth, and the meaning of “epigenetics” has evolved from a general notion to a more specific term; epigenetics is now defined as the study of mechanisms that result in developmental changes in gene activities and effects that are not mediated by altering DNA sequences (2). These changes can be at different cellular levels and include DNA modifications, histone post-translational modifications, RNA-mediated epigenetic regulation, and cellular epigenetic inheritance (2-5).

1.1.2 Chromatin structure and function

One major difference between eukaryotic cells and prokaryotes is that eukaryotic genomic DNA is enclosed within a nuclear membrane in a structure termed nucleus. In the nucleus, DNA is highly condensed and packed with proteins to form a compact structure called chromatin (6, 7). Based on gene transcriptional activity, chromatin exists

in two global forms, heterochromatin and euchromatin (8). Heterochromatin is a densely packed form of DNA that is associated with gene repression and silencing, whereas euchromatin is a loosely packed form of DNA that is associated with active transcription (8). The function of chromatin was originally believed to be packing DNA in the nucleus. Later, it was found that chromatin also plays important roles in preventing DNA damage and regulating gene expression (9).

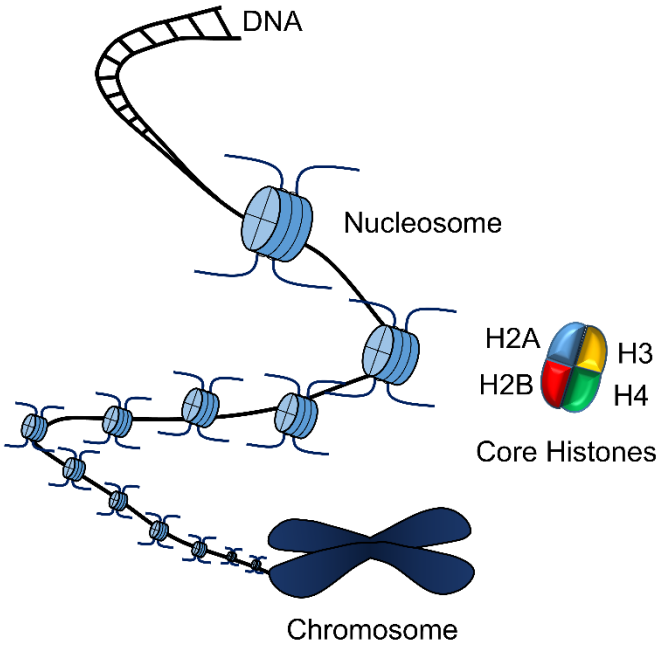


Figure 1.1 Chromatin

In the nucleus of eukaryotic cell, 146 base pairs of DNA are wrapped around a histone octamer, form a fundamental repeating unit called nucleosome. The histone octamer is composed of two copies of each core histone protein (H2A, H2B, H3, and H4).

In the eukaryotic cell, chromatin is composed of a fundamental repeating unit called the nucleosome, where 146 base pairs of DNA are wrapped around a histone octamer (7). The histone octamer is composed of two copies of each core histone protein

(H2A, H2B, H3, and H4) (8, 10). In the histone octamer, H3 and H4 form a $(H3\text{-}H4)_2$ tetramer, and H2A and H2B form two heterodimers (8, 10). It has been suggested that nucleosome assembly occurs through the deposition of $(H3\text{-}H4)_2$ tetramer onto the DNA followed by the rapid incorporation of two H2A-H2B dimers (11, 12). The histone octamers are linked by an interaction between H2B and H4 (8, 10). In addition to the core histones, histone H1 (or H5 in avian species) serves as a linker histone, which is essential for higher-order chromatin structures (8). To form a higher-order chromatin structure, the repeating structure of the nucleosome forms a 30-nm condensed chromatin fiber (8). The chromatin fibers are further packed into structures called chromosomes, which are visible during mitosis and meiosis (9).

The core histones have two distinct regions: the highly ordered histone globular fold regions and the mostly unstructured N- or C-terminal histone tails (10, 13, 14). Both regions have important roles in nucleosome structure and stability. The histone fold regions are necessary for the interactions within histones that allow for histone octamer formation and for the interactions between histones and DNA (10, 13). The unstructured N-terminal tails of both H3 and H2B pass through the DNA superhelix channels and interact with neighboring nucleosomes, thus playing an important role in nucleosome stability and higher-order chromatin structure (10, 14-16).

1.2 Chromatin Modulation

1.2.1 ATP-dependent Chromatin Remodeler

An ATP-dependent chromatin remodeler can regulate chromatin through several mechanisms, including sliding the nucleosomes along DNA, disassembling the histone-DNA complex, evicting the nucleosome, and replacing histones with histone variants (17-19).

There are four classes of ATP-dependent chromatin remodelers. The common feature of these four categories is that they all share a homologous ATPase catalytic domain, which functions in DNA binding and ATP hydrolysis (20, 21). The differences in these four classes are found in the other domains, which have specialized functions (20, 21). The first class of ATP-dependent chromatin remodelers is the SWI/SNF family, which contains a bromodomain (20, 21). The members of the SWI/SNF family interact with acetyl groups on the lysine residues of histone tails through the bromodomain (20, 21). The members of this family mainly promote transcription by sliding nucleosomes along the DNA to enhance the interaction between the transcription factors and the DNA (20, 21). The second class is the ISW1 family, in which the enzymes function in regulating the space between nucleosomes (20, 21). The enzymes in this family usually function in chromosome assembly and gene repression (20, 21). The third class is the CHD family, in which the enzymes are characterized by the presence of a chromodomain (21). The chromodomain has been shown to interact with methylated histone lysine residues, and the CHD family has been shown to promote transcription by sliding or removing nucleosomes (4, 21). One exception is the Mi-2/NuRD complex, which represses gene expression by interacting with histone deacetylases and the H3K4

demethylase (22, 23). The last class is the INO80 family, in which the enzymes have diverse functions, including responding to DNA damage and regulating transcription (20, 21).

1.2.2 Histone Post-Translational Modifications

One of the epigenetic mechanisms by which chromatin can be organized into active or repressed expression modes is the post-translational modification of core histones. The N- and C- terminal tails of the core histones are rich in lysine (K), arginine (R), and serine (S) residues, all of which can be chemically modified. The post-translational modifications of methylation and acetylation were initially discovered on histones in the 1960s (24-28). Since then, over 100 distinct sites on histones that can be chemically modified have been identified (27, 29). These sites are mostly found on histone tails but can also be located in the histone fold domain (27, 29). The most well-known histone modifications include lysine or arginine methylation, lysine acetylation, serine or threonine phosphorylation, lysine ubiquitination, and lysine sumoylation (26, 27, 29). The post-translational modifications to histones can regulate chromatin structure by changing the chemical composition or charge or by serving as docking sites for other effector proteins (29). The histone post-translational modifications have important roles in chromatin structure, gene regulation, gene silencing, the DNA damage response, and cell cycle regulation (29).

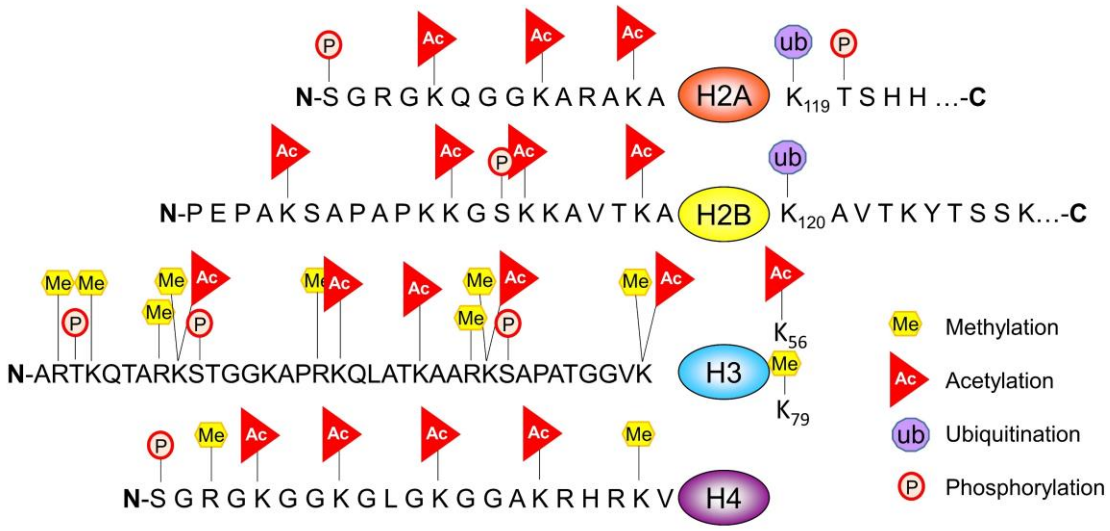


Figure 1.2 Histone post-translational modifications

The histone tails are rich in lysine, arginine, and serine residues, all of which can be chemically modified. The most well-known modifications include methylation, acetylation, phosphorylation, and ubiquitination. The post-translational modifications to histones can regulate chromatin structure through changing the chemical composition or charge or through serving as docking sites for other effector proteins

1.2.2.1 Histone Acetylation, Phosphorylation, and Ubiquitination

Histone acetylation, which is associated with active transcription, was the first histone post-translational modification identified and is the most studied (30-32). The acetylation of histones was first discovered in 1964 (28, 33). Later, histone acetyltransferases (HATs), which transfer an acetyl group from the cofactor acetyl-CoA to the lysine residues of histones (27, 34-36), were identified. Histone acetylation neutralizes the positive charges of the lysine residues on histone tails and, in doing so, decreases the interaction of the histone N-terminal tails with the negatively charged phosphate groups of the DNA (27, 37, 38). Consequently, the chromatin forms a more

relaxed structure, which exposes the DNA to transcription factors and RNA polymerases (27, 31, 38, 39). In addition, acetylated histones serve as docking sites for other functional proteins (40, 41). Bromodomain-containing proteins, including chromatin remodelers and transcription factors, bind to acetylated histones and modulate chromatin (27, 41, 42). The acetyl groups on histones can be removed by histone deacetylases (HDACs). Histone deacetylation results in a compact and condensed chromatin and consequently represses gene expression. There are four classes of HDACs that are based on functional similarity and sequence homology. Classes I, II, and IV share zinc-dependent active sites that can be inhibited by trichostatin A, while Class III HDACs are NAD^+-dependent proteins that are not inhibited by trichostatin A (34, 43). The yeast homologues of the Class I, II, and III families are Rpd3, Hda1, and Sir2, respectively (44). Class IV only contains one member, HDAC11, which does not have a homolog in yeast (44).

Histone phosphorylation plays important roles in the response to genotoxic stress, the regulation of gene transcription, and the regulation of chromatin structure during apoptosis and cell division (45). The phosphorylation of H2A.X S139 in mammalian cells (or H2A S129 in yeast) plays a crucial role in DNA double-strand break repair in all of the phases of the cell cycle and DNA damage response, including replication-coupled DNA repair, homologous recombination (HR), and non-homologous end joining (NHEJ) (45-48). There are several other phosphorylation sites on histone residues that are related to the DNA damage or stress response. For example, histone variant H2A.X tyrosine (Y) 142 is phosphorylated in a DNA damage-dependent manner; and histone H4 S1 phosphorylation is triggered by DNA damage when yeast cells are exposed to UV light

and methyl methanesulfonate (MMS) (45, 49). In addition to the DNA damage response, the phosphorylation of histones also plays a role in crosstalk with other histone modifications, which allows for the proper control of chromatin remodeling and the regulation of gene expression (45). For example, histone H3 S10, threonine (T) 11, and S28 phosphorylation have been shown to be associated with transcriptional activation through Gcn5-dependent acetylation at H3 K9 and K14 (45, 50-52). Additionally, histone phosphorylation plays a role in chromatin compaction during mitosis and meiosis (45). For example, the phosphorylation of histone H3 T3, S10, T11, and S28 play a role in chromosome condensation and segregation (45, 52).

Ubiquitination is a dynamic process that adds a 78-amino acid (8.5-kDa) ubiquitin protein onto a lysine residue of target protein (53, 54). This process requires three steps: ubiquitin is first activated by E1, conjugated to the ubiquitin-conjugating enzyme E2, and finally transferred to the target lysine residue of the substrate with the help of a ubiquitin-protein isopeptide ligase (E3) (53, 54). A substrate can be polyubiquitinated or monoubiquitinated. Polyubiquitination usually results in the substrate being targeted for 26S proteasome degradation, while monoubiquitination primarily marks the substrate for a special function (53, 54). Chromosomes are monoubiquitinated at histones H2A and H2B (55). In higher eukaryotes, histone H2A and histone variant H2A.Z have been identified as targets for ubiquitination, but their ubiquitination is not detected in *Saccharomyces cerevisiae* (56). H2A ubiquitination is related to the repression of gene expression through the regulation of chromatin structure and the repression of histone H3K4 methylation (56). The monoubiquitination of H2B occurs on yeast H2B K123 (K120 in mammalian cells) and is mostly associated with active transcription since the

large size of ubiquitin switches the higher-order structure of chromatin to a more open conformation (55). H2B monoubiquitination is catalyzed by Rad6 (E2) and Bre1 (E3) in yeast, and H2B monoubiquitination is involved in the crosstalk between H3 K4 and K79 methylation (57-59). H2B ubiquitination plays a crucial role in gene transcription, gene silencing, and DNA damage repair during various cellular processes, including the cell cycles, apoptosis, development, and tumorigenesis (57-60).

1.2.2.2 Histone Methylation

Histone methylation was discovered and identified as a stable modification on histones in 1964 (33, 61). Histone methylation is carried out by transferring methyl groups from the cofactor S-adenosyl-L-methionine (SAM) to lysine, arginine, or glutamine residues on the histone (61-64).

Of the three amino acids that can be methylated, histone lysine methylation has been the most studied (65). In human cells, histone lysine residues can be methylated in three states (mono-, di-, or trimethylation), and the major sites of methylation are K4, K9, K27, K36, and K79 on histone H3 and K20 on histone H4 (64, 66). Methylation on H3K4, K36, and K79 is mainly associated with active transcription, and methylation on H3K9 and K27 and H4K20 is primarily associated with gene repression and silencing (37, 63, 64, 66). In *S. cerevisiae*, the major methylation sites are on histone H3K4, K36, and K79 and are primarily associated with active transcription (64, 66). Intriguingly, methylation on histones H3K9 and K27, which is mainly associated with gene repression

and silencing, has not been identified in *S. cerevisiae* (66). One recent study identified histone H4K20 monomethylation in *S. cerevisiae* and found that it functioned in subtelomeric silencing (67). However, the histone methyltransferase (HMT) for this site has not been identified in *S. cerevisiae*, and there have been no follow-up studies (67). Recent studies have identified Set5 as the a HMT responsible for the monomethylation of histones H4K5, K8, and K12 in *S. cerevisiae* (68, 69).

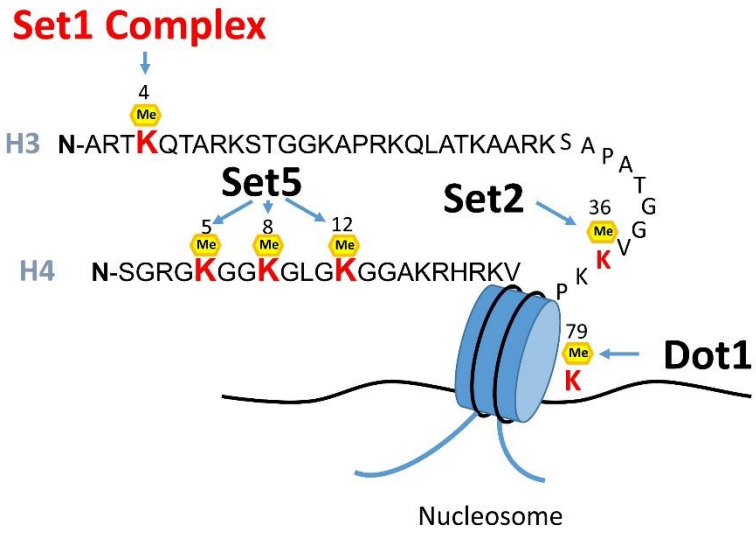


Figure 1.3 Histone H3 and H4 methylation sites in Saccharomyces cerevisiae

In Saccharomyces cerevisiae, Set1 complex, Set2, and Dot1 methylate H3K4, K36, K79 respectively. All three markers have been associated with positive transcription. Recent study shows Set5 can monomethylate H4 K5, K8, and K12.

Histone methylation is catalyzed by histone lysine methyltransferases (KMTs) using SAM as a cofactor (6, 70). Most KMTs contain a SET domain in their catalytic core, except for yeast Dot1 (human DOT1L), the KMT for histone H3K79 (64, 70). The catalytic domain of Dot1 shares protein sequence similarity to histone arginine

methyltransferases, even though Dot1 has not been identified as a histone arginine methyltransferase (71, 72).

Histone lysine methylation is both a stable and dynamic modification, and the methyl groups can be removed from lysine by histone lysine demethylase (HDM) (73). There are two families of HDM that have been identified, the LSD family and the JmjC family (74). The LSD family demethylates histones through a FAD (flavin adenine dinucleotide)-dependent amine oxidase reaction (74). The LSD homologues are not found in *S. cerevisiae* (75). The JmjC domain family demethylates histones through a 2OG-Fe(II)-dependent dioxygenase reaction, and this motif is conserved from yeast to humans (74). In *S. cerevisiae*, Jhd2, a JmjC domain-containing demethylase, can remove histone lysine methylation on H3K4 (76-78).

The additional methyl groups on histones do not change the positive charge of lysine or arginine (65). Rather, methylated histones mainly serve as docking sites for protein-protein interactions that regulate chromatin structure and gene expression, such as those encoding chromatin remodelers, enzymes required for the post-translation modification of histones, and transcription factors (65, 79-82). The specific binding domains for histone lysine methylation include WD40 domains, chromodomains, Tudor domains, and PHD fingers (81, 83). For example, human HP1, a chromodomain-containing protein, is reported to bind methylated H3K9 and function in DNA methylation and RNA-mediated silencing (84). In addition, yeast Set3, a PHD finger- and SET domain-containing protein, has been shown to interact with H3K4 dimethyl groups and function in histone deacetylation (85-87).

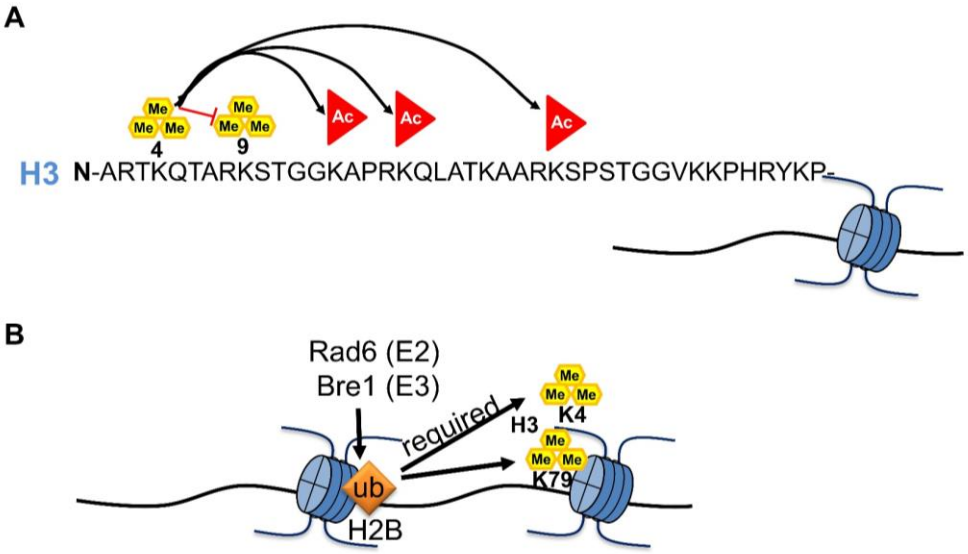


Figure 1.4 H3K4 trimethylation cross-talks with other histone modifications

(A) H3K4 trimethylation promote H3 acetylation and inhibit H3K9 methylation. (B) The Rad6 and Bre1 mediated Histone H2B monoubiquitination is required for H3K4 and K79 methylation.

Histone arginine methylation can be associated with either transcriptional activation or repression and is mediated by protein arginine methyltransferases (PRMTs) that deposit methyl groups on histone H3R2, R8, R17, and R26 and H4R3 (37, 88, 89). The PRMTs catalyze three forms of arginine methylation: monomethylation, asymmetric dimethylation, and symmetric dimethylation (88). Recent studies have shown that JMJD6, a JmjC domain-containing protein, is a putative demethylase of histone H3R2 and H4R3 dimethylation (88, 90). Additionally, there is a PADI4-mediated deamination reaction that converts an arginine to citrulline to prevent the methylation (91). Histone glutamine methylation was recently identified on H2A glutamine (Q) 104 in humans (and Q105 in yeast) and was found to be exclusively enriched at the 35S ribosomal DNA (rDNA) locus (62). In yeast, H2A glutamine methylation is mediated by Nop1 to prevent

the recruitment of FACT to the 35S rDNA locus and results in a more open chromatin conformation (62).

1.2.2.3 Histone H3 Lysine 4 Methylation

Histone H3 Lysine 4 (H3K4) methylation is one of the most well-studied post-translational modifications. H3K4 methylation is conserved from yeast to humans and is generally considered to be a hallmark of active gene transcription (29, 92). Genome-wide studies have found that the pattern of H3K4 trimethylation enrichment is strongly correlated with actively transcribed regions from yeast to humans. In yeast, H3K4 mono-, di-, and trimethylation are mainly enriched on the open reading frames (ORFs) of active genes (29, 92). H3K4 trimethylation is predominately enriched at the transcriptional start sites and at the 5' ends of the ORFs of active genes (29, 92). H3K4 dimethylation is mainly enriched in the middle of the ORFs, and monomethylation is enriched at the 3' ends of active genes (29, 92).

H3K4 methylation sites serve as molecular docking sites for the recruitment of effector proteins. (81, 93, 94) H3K4 methylation either activates or represses transcription depending on the function of the effector protein (81, 94). For example, CHD1, a chromodomain-containing and H3K4 di- and trimethylation-interacting protein, alters gene expression by recruiting SAGA HAT complexes (95, 96). Alternatively, Set3 interacts with H3K4 dimethylation and functions in histone deacetylation and subsequent gene repression (85-87).

As a part of the "histone code," histone H3K4 methylation is involved in crosstalk with other histone modifications (63, 81, 94). The enrichment of H3K4 methylation is

correlated with the location of histone acetylation (64, 87). H2BK123 monoubiquitination (H2Bub) in *S. cerevisiae* is related to H3K4 methylation through a trans-tail mechanism (57, 59). In contrast, the location of H3K4 methylation could prevent the presence of other modifications, such as H3K9 methylation on the histone tails (64, 97).

1.3 Set1: the SET-domain Containing Protein

1.3.1 SET-domain Containing Proteins

The SET domain is an evolutionarily-conserved motif of approximately 130 amino acids (98). It was named after the three original *Drosophila* proteins, Su(var)3-9, enhancer-of-zeste, and Trithorax (99). The SET domain mainly functions as a catalytic domain in most KMTs (except for Dot1) and acts by transferring a methyl group from SAM to the ε-amines of lysine residues on histone tails (100). The SET domain is flanked on either side by less-conserved n-SET (or pre-SET) and post-SET domains. The n-SET domain stabilizes the SET domain, and the disordered post-SET domain has been shown to be associated with the SAM cofactor and histone tails (101, 102). The SET domain forms a specific fold pseudoknot structure and creates an active site in which a zinc ion is coordinated to three conserved cysteine residues in the post-SET domain (101, 102). In *S. cerevisiae*, there are seven SET domain-containing proteins, named Set1 to 6 and Pkm4 (103, 104). Pkm4 was originally named Set7 and was later identified as a ribosomal lysine methyltransferase that targets lysine 55 of Rpl42ap and Rpl42bp (104). Set1, Set2, and Set5 have been identified as histone methyltransferases that target H3K4; H3K36; and H4K5, K8, and K12, respectively (68, 69, 103, 105). Although Set3 also contains a

SET domain, the methyltransferase activity of Set3 has not been identified. Instead, Set3 has been shown to bind to H3K4 dimethylation sites and function as a subunit of the histone deacetylase complex (87). The functions of Set4 and Set5 have not been identified.

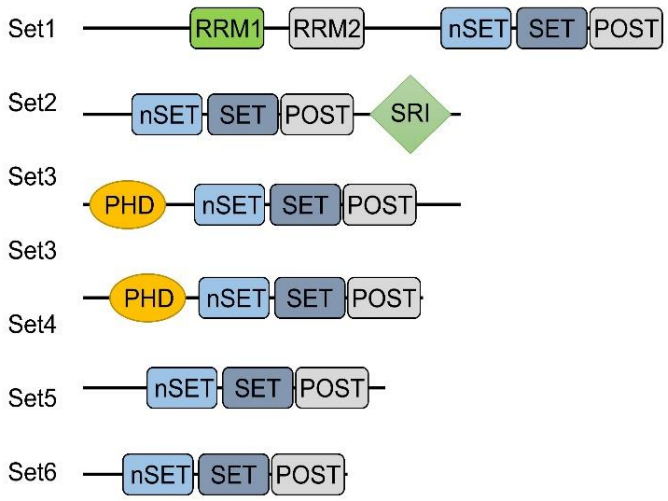


Figure 1.5 Schematic representation Domain structures of SET domain containing proteins in yeast

1.3.2 Set1: the Histone H3K4 Methyltransferase in *Saccharomyces cerevisiae*

The first H3K4 methyltransferase identified was Set1 in *Saccharomyces cerevisiae.* After analyzing individual deletion strains for each of the seven SET domain-containing proteins in *Saccharomyces cerevisiae*, Set1 was found to be the only protein responsible for H3K4 methylation (103). This and the following results demonstrated that Set1 is the sole H3K4 methyltransferase that is responsible for H3K4 mono-, di-, and trimethylation in *S. cerevisiae* (103, 106-108). Unlike other yeast HMTs, such as Set2 and Dot1, Set1 must be present in a multi-protein complex for it to be active *in vitro* and *in vivo* (92, 106). In *S. cerevisiae*, the Set1 complex (Set1C) or Complex of Proteins Associated with Set1 (COMPASS) consists of eight members: Set1, Bre2, Sdc1, Swd1,

Swd2, Swd3, Spp1, and Shg1 (106, 107, 109). The Bre2, Sdc1, Swd1, Swd3, and Spp1 subunits are required for various degrees of proper H3K4 methylation, and Swd1, Swd2, Swd3, and Spp1 have been shown to be important for regulating Set1 protein levels, suggesting that these subunits may contribute to Set1 protein stability (106, 107, 109, 110). A recent study showed that Set1 methylates a non-histone target, the kinetochore protein Dam1 (111, 112).

Following the identification of yeast Set1, the human Mixed Lineage Leukemia family (MLL1 through 4) and human SET1A and SET1B were identified as yeast Set1 homologues. All of the human H3K4 methyltransferases form complexes and share most of the core subunits that are homologous to subunits of the yeast Set1 complex (82). The core subunits consist of ASH2 or ASH2L (Bre2 homolog), DPY30 (Sdc1 homolog), RBBP5 (Swd1 homolog), WDR82 (Swd2 homolog), WDR5 (Swd3 homolog), and CFP1 or CGBP (Spp1 homolog) (82). WDR82 has only been identified in human SET1 complexes and not in MLL complexes (82). Similar to that observed in yeast, disruptions of the human MLL(1-4)/SET1A/B complex subunits WDR5, RBBP5, and ASH2L reduce H3K4 trimethylation. These results suggest that studying yeast Set1C can help us understand how human H3K4 methyltransferase complexes function.

As the catalytic core of Set1C, Set1, a 1080-amino acid protein, consists of two putative N-terminal RNA recognition motifs (RRM1 and 2, amino acids 210-335 and 415-568, respectively), an n-SET domain (787-935 a.a.), a SET domain (936-1063 a.a.), and a post-SET domain (1064-1080 a.a.) (106, 113, 114). The SET domain is a catalytic domain and is necessary for histone methyltransferase activity (113). A deletion or mutation in the n-SET or post-SET domain also results in the loss of all histone H3K4

methylation (113). The n-SET domain contains a motif of basic patch residues that is necessary for the association of Set1 with Swd1 (110). Interestingly, a Set1 truncation mutant containing the n-SET domain, SET domain, and post-SET domain can associate with the Set1C subunits, Spp1, Bre2, Sdc1, Swd1, and Swd3 and is the minimal unit necessary for global H3K4 trimethylation *in vivo* and *in vitro* (115, 116). However, the truncated Set1 was expressed at a higher level than endogenous Set1, thus, global H3K4 trimethylation *in vivo* may be due to a gain of function. Set1 contains two putative RRM domains in the N-terminus. Interestingly, human Set1A and Set1B also contain a conserved RRM domain, while MLL1-4 does not (117). It was originally thought that the function of the RRM domain is to bind RNA, but it was later discovered that some RRM domains can promote protein-protein interactions (113). The deletion of RRM1 of Set1 results in a loss of H3K4 trimethylation but has little effect on di- and monomethylation (113). One study showed that the RRM1 and RRM2 domains together can non-specifically interact with RNA, whereas the RRM1 domain alone cannot (118). Human SET1A and SET1B contain only one RRM domain, and one study showed that the N-terminus of human SET1A that contains the RRM domain can interact with Ser5-phosphorylated CTD peptides with the help of Wdr82 (Swd2 homolog in humans) *in vitro* (117). Further studies are needed to elucidate the function of RRM domains in yeast.

In Set1C, Bre2 and Sdc1 form a heterodimer and have related functions. The deletion of either *BRE2* or *SDC1* results in a loss of global H3K4 trimethylation and a decrease in mono- and dimethylation (106, 107, 109, 110). Bre2, a 505-amino acid protein, is the second largest subunit in Set1C. Sdc1, a 17-amino acid protein, is the

second smallest subunit in Set1C (106). Bre2 is composed of two domains, an N-terminal SPRY domain and a C-terminal SDI domain (119). The function of the SPRY domain in yeast is unknown, but a study has shown that the SPRY domain in human ASH2L (Bre2 homolog) is required for the interaction of ASH2L with human RBBP5 (Swd1 homolog) and the cofactor SAM. (120, 121). The SDI domain of Bre2 has been shown to interact with the Dpy-30 domain of Sdc1 (119).

Set1C has three WD40 repeat-containing subunits, Swd1 (426 a.a.), Swd2 (329 a.a.), and Swd3 (315 a.a.). The WD40 domain has been shown to form a β-propeller structure that is involved in protein-protein interactions (114, 122). The deletion of either *SWD1* or *SWD3* causes a loss of all levels of H3K4 trimethylation and a significant decrease in Set1 protein levels (106, 107, 109, 110). A possible explanation for this is that Swd1 and Swd3 form a heteromeric complex independent of Set1, and the deletion of either one will therefore result in Set1 losing its interaction with the other (106, 110, 114). In addition to the WD40 domain, the C-terminus of Swd1 contains two acidic patches that are necessary for its association with the basic patches within the n-SET domain of Set1 (110). These two acidic patches are also required for H3K4 methylation and Set1 protein stability (110). The function of Swd3 in Set1C activity has not been identified. Studies have shown that the human WDR5 (Swd3 homolog) interacts with the K4 residue on the H3 tail (123). Swd2 is an essential gene and is also a member of the APT cleavage and polyadenylation factor complex (106). Conditional mutations suggest that Swd2 is needed for proper H3K4 trimethylation (124).

Spp1 (353 a.a.) is a PHD finger-containing protein. The deletion of *SPP1* results in a modest reduction in both H3K4 trimethylation and Set1 protein levels (106, 107,

109, 110, 124, 125). An *in vitro* peptide-protein binding assay showed that Spp1 can interact with histone H3K4 di- and trimethylated peptides through its PHD finger and that Spp1 can interact with DNA via its C-terminus (126, 127). Spp1 is associated with the n-SET domain of Set1 (115). However, the biological function of Spp1 needs further characterization.

Shg1 (142 a.a.) is the smallest subunit of Set1C. The function of Shg1 is unknown, and there is no human homolog of Shg1. The deletion of *SHG1* does not affect H3K4 methylation (110). Shg1 directly associates with Set1 (356-568 a.a.), which is a region of Set1 RRM2. However, the biological function of this interaction is still unknown (115).

Before Set1 enzymatic activity was identified, studies in yeast showed that Set1 is needed for normal yeast morphology and growth and for gene silencing at rDNA, telomeres, and mating-type loci (103, 107, 108, 113, 128). However, a later-proposed mechanism suggested that H3K4 trimethylation prevents Sir2 HDAC complex from spreading at these silencing regions; without H3K4 trimethylation, the Sir proteins titrated away from silent chromatin, causing loss of silencing (129, 130). The most recent studies have shown that the deletion of *SET1* results in growth defects during the response to various types of stress, such as DNA damage and oxidative stress (131, 132). Set1 also regulates gene expression in the ergosterol pathway (133). The antifungal drug Brefeldin A (BFA) targets the yeast ergosterol pathway, and the deletion of *SET1* or Set1 complex subunits results in sensitivity to BFA at different degrees (133). Further studies are needed to understand the other biologically relevant roles of Set1 and H3K4 methylation in yeast.

1.4 RNA polymerase II and Set1

In eukaryotic cells, the process of gene transcription is executed by three DNA-dependent RNA polymerases (134). RNA polymerase II (RNAPII) is the most studied and functions in the transcription of pre-mRNA and small nuclear RNA (134). RNAPII is conserved in eukaryotes and has a total molecular weight of approximately 550 kDa (134). RNAPII is composed of 10-12 subunits (12 subunits in yeast and humans) (134, 135). RNAPII is composed of a core enzyme of 10 subunits and a peripheral Rpb4/7 subcomplex (135). The Rpb4/7 subcomplex can be dissociated from purified RNAPII and has been shown to stimulate the selective initiation of RNAPII (134, 135).

1.4.1 The Phosphorylation on CTD of RNA polymerase II

The C-terminal domain (CTD) of RNAPII is on its largest subunit, Rpb1, which is composed of heptapeptide repeats of (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7). There are 26 repeats in yeast and 52 repeats in mammals (136). The CTD is essential for viability, although the number of repeats can be reduced (137-139). The CTD is also essential for cell growth. Yeast require at least 10 CTD repeats for viability, and a yeast strain containing only 10-12 CTD repeats has a strong temperature-sensitive growth defect (137-139). In the pre-initiation complex, RNAII with an unphosphorylated CTD is loaded on chromatin (140). The CTD is then modified at multiple sites, mostly by phosphorylation, so that the CTD codes will be presented at the different stages of transcription (140). The different CTD codes will recruit stage-specific factors that play important roles in transcription and CTD modification (140). The most studied CTD modifications are phosphorylation at Ser2, 5, and 7 (140). Phosphorylated Ser5 and Ser7

occur at the transcription start site in the 5’ region, Ser5 phosphorylation is replaced by Ser2 phosphorylation during transcription elongation so that Ser2 phosphorylation is enriched at the 3’ ends of the ORF, and Ser7 phosphorylation remains at high levels across the ORF (140). At the transcription start site, Ser5 is phosphorylated by Kin28 (CDK7 in humans) and Srb10 (CDK8 in humans), and Rtr1 is the phosphatase responsible for the removal of Ser5 phosphorylation shortly after initiation (140). Kin28 also phosphorylates Ser7 at the transcriptional start site, and the kinase Bur1 is responsible for maintaining Ser7 phosphorylation at the 3’ end (140). Throughout the elongation process, Ctk1 is the primary kinase for Ser2 phosphorylation. Bur1 (CDK9 in humans) is also a Ser5-P-dependent Ser2 kinase, and Ser2-P and Ser5-P can co-exist on the CTD, and this is important for recruiting elongation regulators, such as Set2 (140, 141). Fcp1 dephosphorylates Ser2-P upon the termination of transcription. The remaining phosphorylated Ser5 and Ser7 sites at the 3’ end are removed by the phosphatase Ssu72 (140).

There are other modification sites on the CTD. Tyr1 is phosphorylated both in human cells and yeast, although the yeast kinase has not been identified (142). Tyr1 is phosphorylated along with Ser2-P, but Tyr1-P is decreased upstream of the poly-A region (142). Tyr1-P has been shown to enhance the binding of elongation factors and to prevent the recruitment of termination factors (140, 142). Thr4 can also be phosphorylated. The Thr4 substitution does not affect yeast growth; however, it is lethal to higher eukaryotes (140, 143). Other types of modifications have either not been shown in yeast or have functions that are not well characterized (140).

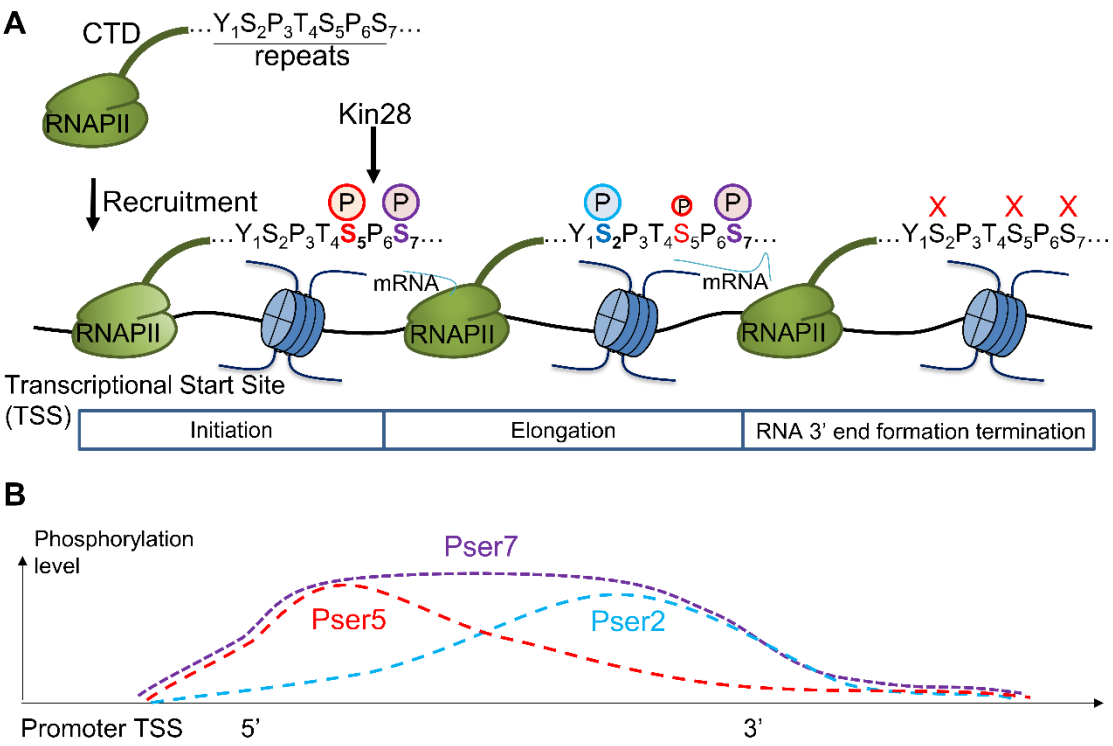


Figure 1.6 Phosphorylation of CTD of RNA Polymerase II (RNAPII) and transcription.

(A) CTD is phosphorylated in different steps of transcription. The largest subunit of RNAPII, Rpb1 containing a flexible C-terminal domain (CTD) which consists of multiple repeat of 7 amino acids (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7). Before RNAPII is loaded on to the chromatin, the serine amino acids are not phosphorylated. After RNAPII recruited to transcriptional start site, Ser5 and Ser7 are phosphorylated, in yeast the kinase is Kin28. When the RNAPII move along the DNA, called elongation, the Ser5 phosphorylation is removed and Ser2 is phosphorylated. And when RNAPII approaching to 3' of gene and forming termination complex, all the serine phosphorylation on the CTD is removed. (B) The pattern of CTD phosphorylation on the open reading frame of active gene. Ser5-P is enriched at TSS and 5' ends, Ser2-P is enriched at 3' ends of ORF, and Ser7-P is enriched over the ORF.

Phosphorylation of the RNAPII CTD functions along with histone modifications to regulate gene transcription. It has been shown that Ser5-P associates with the Set1 complex (140). Thus, Set1 and H3K4 trimethylation were found by chromatin

immunoprecipitation (ChIP) analysis to be highly enriched at actively transcribed genes near the 5’ ends of coding regions (140). During elongation, Ser5-P and Ser2-P both associate with Set2 and methylate H3K36. H3K36 trimethylation, which functions in transcription termination and prevents cryptic transcription, is primarily enriched at the 3’ end (140).

1.4.2 RNA Polymerase II Association Factor (Paf1) Complex and Set1

The RNA polymerase II association factor complex I (Paf1C) is composed of five subunits, Paf1, Ctr9, Cdc73, Rtf1, and Leo1 in yeast and is conserved from yeast to humans (144-146). The Paf1 complex plays important roles in multiple steps of transcription, including transcription elongation, the transcription-related modification of histones, and the termination of mRNA synthesis (147). Paf1C is localized to the open reading frames of active genes and genetically and physically interacts with RNAPII and elongation factors, such as Spt5 (147). Mutations in Paf1C lead to sensitivity to the base analog 6-azauracil (6-AU) and mycophenolic acid, both of which affect transcription elongation (147). Additionally, the Paf1 complex regulates the recruitment of the E2 ubiquitin-conjugating enzyme Rad6 and the E3 ubiquitin ligase Bre1 to active genes to monoubiquitinate H2B, which is a requirement for H3K4 and H3K79 trimethylation by Set1C and Dot1 (148). The trimethylation of H3K36 also requires Paf1C, which plays an important role in the repression of cryptic transcription (147). Paf1C is also required for proper 3’ end processing and the termination of mRNA and snoRNA transcription (149).

The Paf1 complex has been shown to genetically interact with the Set1 complex, as double-deletion of the Paf1 and Set1 complex subunits causes synthetic lethality in yeast

(150) The current model suggests that the Paf1 complex recruits the Set1 complex to interact with the Ser5-phosphorylated CTD of RNAPII (150-152). However, this model is based on the Paf1 complex-dependent interaction between RNAII and Bre2, a Set1 complex subunit, rather than on the detection of a direct interaction between Set1 and RNAPII (150-152). Without Bre2, the yeast still has partial global H3K4 monomethylation and dimethylation (110). Further investigations are needed to elucidate the relationship between the Paf1 and Set1 complexes.

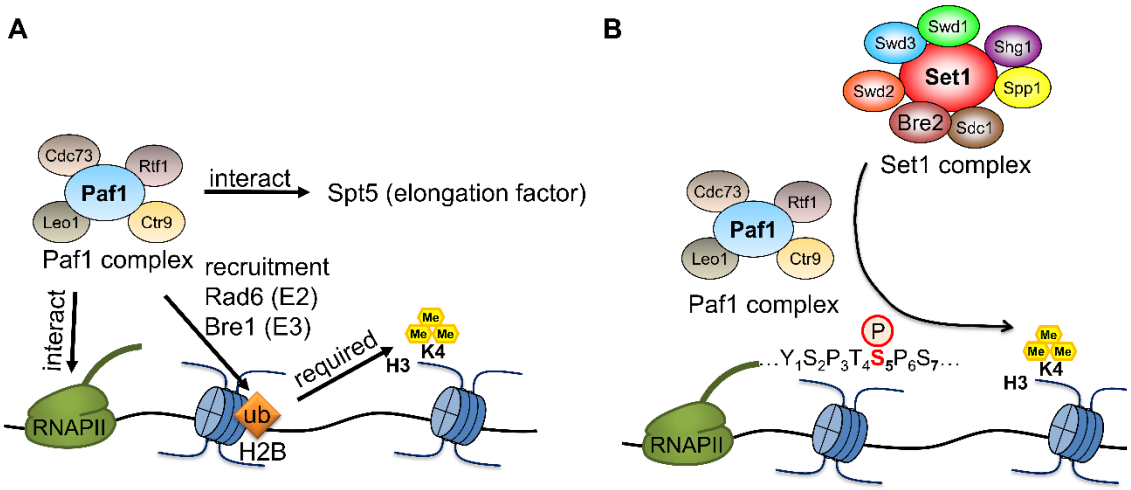


Figure 1.7 The model of Paf1 Complex is required for H3K4 methylation.

(A) The Paf1 complex plays important roles in multiple steps of transcription. It interacts with RNAPII and elongation factors for example Spt5. It also required for Rad6 and Bre1 mediated H2B ubiquitination, which is required for H3K4 trimethylation. (B) The current model suggests Set1 complex is recruited to Ser5-P CTD of RNAPII through Paf1 complex.

CHAPTER 2. THE ROLE OF THE SET1 RNA POLYMERASE II INTERACTING MOTIF (SRIM) IN SET1 RECRUITMENT AND HISTONE H3K4 METHYLATION

2.1 Declaration of collaborative work

In this study, all the experiments and analysis were performed by Yueping Zhang.

2.2 Introduction

In the eukaryotic cell, gene expression is highly regulated by epigenetic modifications, including methylation, acetylation, phosphorylation, and ubiquitination. Among these modifications, histone H3 lysine 4 (H3K4) methylation has been associated with active gene transcription (64, 92). In *S. cerevisiae,* Set1 is the sole histone methyltransferase responsible for H3K4 mono-, di-, and trimethylation (103, 106-108). H3K4 trimethylation (H3K4me3) is primarily enriched at the transcriptional start site and 5’ end of the open reading frame (ORF), H3K4 dimethylation (H3K4me2) is mainly enriched in the middle of the ORF and H3K4 monomethylation (H3K4me1) is enriched in 3’ of the ORF(153-155). Recent studies into the function of these methylation events show they act as binding sites to recruit effector proteins (81, 93, 94). Depending on the function of the effector protein, they can elicit a response to help activate or repress gene transcription (85-87, 95, 96).

Unlike other yeast histone methyltransferases such as Set2 and Dot1, Set1 functions in a complex with seven other subunits (Bre2, Swd1, Spp1, Swd2, Swd3, Sdc1, and Shg1), together called Set1 complex (Set1C or COMPASS) (106, 107, 109). The Set1C subunits are important for Set1 catalytic activity *in vitro* and *in vivo* (92, 106). The human homologs of Set1 include the Mixed Lineage Leukemia family (MLL1-4) and the human SET1 family (Setd1A and Setd1B). All of the human H3K4 methyltransferases form complexes and share most of the core subunits that are homologous to subunits of the yeast Set1 complex (82, 152). As the catalytic core of Set1C, Set1, a 1080-amino acid protein, consists of two putative N-terminal RNA recognition motifs (RRM1 and 2, amino acids 210-335 and 415-568, respectively), an n-SET domain (797-933 a.a.), a SET domain (933-1063 a.a.), and a post-SET domain (1064-1080 a.a.) (106, 113, 114). The function of RRM domains are generally thought to bind RNA, but it was later discovered that some RRM domains can participate in protein-protein interactions (113). The RRM1 domain of Set1 is required for H3K4 trimethylation and Set1 stability (113). The RRM2 domain of Set1 has been shown to be potentially associated with RNA (118). Interestingly, human Set1A and Set1B also contain only one conserved RRM domain, while MLL1-4 does not (117). The study showed that the N-terminus of human SET1A that contains the RRM domain can interact with Ser5-phosphorylated (Ser5-P) CTD peptides with the help of Wdr82 (Swd2 homolog in humans) *in vitro* (117). MLL1 and MLL2 have also been shown to interact with Ser5-P CTD. (156, 157)

The C terminal domain (CTD) of RNAPII is on its largest subunit, Rpb1, which is composed of heptapeptide (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) repeats, and there are

26 repeats in yeast and 52 repeats in mammals (158). During transcription initiation and elongation, RNAPII successively recruits stage-specific factors, by Serine phosphorylation on the CTD (159). In yeast, Serine 2, 5, 7 can be phosphorylated. Ser5 phosphorylated (Ser5-P) CTD is enriched at the transcription start site 5' region; and during transcription elongation, Ser5 is dephosphorylated and Ser2 is phosphorylated; Ser7 phosphorylation is enriched across the ORF (160). Set2 directly interacts with CTD phosphorylated at both Ser2 and Ser5 of RNAPII through its Set2 RNA polymerase II interaction (SRI) domain, which help to recruit Set2 to middle and 3' end of ORF to deposit H3K36 trimethylation (141). Set1 has been shown to interact with Ser5-P CTD of RNAPII, however, it is still unclear how this interaction occurs and whether Ser5-P CTD of RNAPII helps the deposition of H3K4 trimethylation on ORF (161). The study of how yeast Set1 associates with RNAPII will help us understand the mechanism of Set1 recruitment and H3K4 methylation deposition.

Paf1C functions in the transcriptional elongation complex, which is composed of five proteins: Paf1, Leo1, Rtf1, cdc73, and Ctr9 (144-146). Paf1C is genetically linked to Set1C, and double deletion of Paf1C subunits and Set1C subunits causes synthetic lethality in yeast (150). Paf1C recruits Rad6 and Bre1 to chromatin to perform monoubiquitination on H2B K123, which is required for H3K4 trimethylation (162). The current model claims that Paf1C recruits Set1C to interact with Ser5-P CTD of RNAPII. Characterizing whether Paf1C affects Set1 and RNAPII containing P-Ser5 CTD interaction can help us understand how Set1 is recruited to chromatin to deposit H3K4 methylation.

2.3 Results

2.3.1 Set1 interacts with RNAPII containing Ser5-P and/or Ser7-P CTD and this interaction is independent of Paf1C subunits and most Set1C subunits

Set1 has been shown to be associated with the Ser5-P CTD of RNAPII (161), however, the specificity of this interaction is still unclear. Thus, we generated an N-terminal 3×FLAG tagged endogenously expressed *SET1* strain. The whole cell extracts of indicated strains were immunoprecipitated using α-FLAG (M2) conjugated magnetic beads and analyzed by Western blot using antibodies against specific Serine phosphorylation on CTD of RNAPII. As shown in Figure 2.1A, Set2, as a positive control, interacts with not only Ser2-P and Ser5-P CTD of RNAPII as previous published, but also Ser7-P CTD of RNAPII, which is a novel discovery (141). In contrast, Set1 interacts RNAPII containing Ser5-P and/or Ser7-P CTD, not Ser2-P. To rule out the possibility that CTD antibodies may cross-react with other Serine phosphorylated forms of CTD, in addition to the antibodies of H5 (α-Ser2P) and H14 (α-Ser5P) used in Figure 2.1A, we also applied another α-Ser5P antibody (3E8) to the blot and the result is similar (163). The set2 mutant without Set2 Rpb1 interacting (SRI) domain is used as a negative control (141).

Previous studies in our lab and others have shown some Set1C subunits are required for proper Set1 functions as histone H3K4 methyltransferase (106, 110, 114, 119). It is possible that Set1 interacts with RNAPII through Set1C subunits. To investigate whether Set1 is required for Set1C subunits interact with RNAPII, all Set1C subunits were C-terminal 3×FLAG tagged in both wild-type and *set1*Δ strains. The whole

cell extracts of these strains were prepared and subjected to α-FLAG (M2) immunoprecipitation assay and analyzed by Western blot using α-Ser5P (H14) and Rabbit α-FLAG antibodies. The result indicated that all Set1C subunits can interact with RNAPII containing Ser5-P CTD, and these interactions are abolished when *SET1* is deleted (Figure 2.1B). The loss of interactions is not due to degradation of subunits without Set1, for the protein levels of each subunits remain the same in the *set1*Δ strains compared to *SET1* strains (Figure 2.1B). These results demonstrate that Set1 is necessary for Set1C associating with RNAPII containing Ser5-P CTD.

After we identified Set1 is required for Set1C interacting with RNAPII, we wanted to determine whether Set1C subunits are required for Set1 to interact with RNAPII. Thus, *bre2*Δ*, swd1*Δ*, swd3*Δ, *spp1*Δ, *sdc1*Δ, and *shg1*Δ were generated in *3×FLAG-SET1* strain. Because *SWD2* is essential for yeast to grow, we did not test the dependency of Swd2 (152). The whole cell extracts of indicated strains were prepared and subjected to α-FLAG (M2) immunoprecipitation assay and analyzed by Western blot using α-Ser5P (H14) and α-FLAG (M2) antibodies. As shown in Figure 2.1C, Set1 still interacts with RNAPII containing Ser5-P CTD in all tested deletion strains. Though the immunoprecipitated Ser5-P CTD levels were decreased in *swd1*Δ*, spp1*Δ, and *swd3*Δ, and *shg1*Δ, the immunoprecipitated Ser5-P CTD levels are correlated with the Set1 protein levels in corresponding strains. This data indicates that the Set1-RNAPII interaction is independent of most Set1C subunits, including Bre2, Swd1, Spp1, Swd3, Sdc1, and Shg1.

Associated with RNAPII and crucial for transcription elongation, Paf1C is composed of five subunits, Paf1, Leo1, Rtf1, cdc73, and Ctr9 (144-146). Current model in the field indicates Set1-RNAPII interaction is mediated by Paf1C (150-152). To determine whether Paf1C is required for Set1 to interact with RNAPII, individual Paf1C subunit deletions, *paf1*Δ, *ctr9*Δ, *cdc73*Δ, *rtf1*Δ, and *leo1*Δ, were generated in *3×FLAG-SET1* strain. The immunoprecipitation assay was performed as described in Figure 1C. To our surprise, Set1 still interacts with RNAPII containing Ser5-P CTD in all individual Paf1C subunit deletion strains (Figure 2.1D). We also determined the global levels of H3K4 mono-, di-, and trimethylation and H2B ubiquitination in our Paf1C subunit deletion strains, and the results are same as published results (Figure 2.1E). Our result indicates that the Set1-RNAPII interaction is independent of Paf1C.

2.3.2 A novel domain in the N terminus of Set1 mediates Set1 and RNAPII interaction.

Given that Set1 and RNAPII interaction is independent of Paf1C subunits and most Set1C subunits, we sought to identify the region on Set1 responsible for this interaction. Yeast *CEN* constructs were generated that encode 3×FLAG tagged full length or truncated Set1, and the Set1 protein is expressed from *SET1* promoter (-400 to -1bp) (Figure 2.2A). The indicated constructs were transformed to *set1*Δ strain, and whole cell extracts of indicated strains were immunoprecipitated as described in Figure 2.1C. Interestingly, as shown in Figure 2.2B, the N terminus of Set1 ($set1_{1\text{-}689}$ and $set1_{1\text{-}375}$) interacts with RNAPII containing Ser5-P CTD, whereas the C terminus of Set1 ($set1_{690\text{-}1080}$ and $set1_{380\text{-}1080}$) fails to. This result indicates that $set1_{1\text{-}375}$ is necessary and sufficient to interact with RNAPII containing Ser5-P CTD. To further define the Set1-RNAPII

interacting sites, Set1 constructs with sequential N-terminal truncations were generated (Figure 2.2C). The constructs were transformed and tested for RNAPII interaction through immunoprecipitation assay as described in Figure 2.1C. As shown in Figure 2.2D, set1$_{102\text{-}1080}$ still interacts with RNAPII containing Ser5-P CTD, while Set1$_{151\text{-}1080}$ and Set1$_{211\text{-}1080}$ fail to. This result indicates that the amino acids 102-151 of Set1 are required for Set1 and RNAPII interaction. Given set1$_{1\text{-}375}$ is necessary and sufficient for Set1-RNAPII interaction and we want to further narrow down the Set1 region that is sufficient for RNAPII binding, more truncations in amino acids 1-375 of Set1 were generated. Although set1$_{1\text{-}375}$ is a stable truncation, the further truncations were found to be poorly expressed. Thus, 2-micron overexpression constructs of set1$_{151\text{-}375}$, set1$_{1\text{-}275,}$ and set1$_{1\text{-}250}$ were generated and immunoprecipitation assay was performed as described in Figure 2.1C. To our surprise, none of these Set1 truncations bind to RNAPII containing Ser5-P CTD (Figure 2.2F). The RRM1 domain localizes at amino acids 230-335, and both set1151-375 and set1211-1080 contain RRM1 and do not interact with RNAPII. Together, these data indicate that RRM1 cannot interact with RNAPII by itself and both RRM1 domain and the uncharacterized region upstream of RRM1 are required for Set1-RNAPII interaction.

Since amino acids 102-151 of Set1 is in uncharacterized region (upstream of RRM1), we performed BLAST analysis of this region and identified two conserved sites: RP (amino acids 133, 134) and KFHYFD (amino acids 148-153) (Figure 2.3A). Interestingly, all the identified proteins are Set1 orthologues in different genera under the same family of *Saccharomycetaceae* (Figure 2.3A), however, these sites cannot be

identified in a different family or in other higher eukaryotes, however, there might be some very distant conservation. To test whether these two sites are required for Set1 and RNAPII interaction, we generated *CEN* plasmids containing 3×FLAG tagged Set1 and set1 point mutants in the conserved region, and all Set1 and set1 mutants were expressed from a *SET1* promoter. Then, the indicated vectors were transformed to the *RPB3-3×HA*, *set1*Δ strain. The whole cell extracts of indicated strains were prepared and subjected to α-FLAG (M2) immunoprecipitation and analyzed by Western blot using α-Ser5P (H14), α-FLAG (M2), and α-HA (12CA5) antibodies. As shown in Figure 2.3B, the Set1-Ser5-P-CTD interaction and Set1-Rpb3 interaction are disrupted by mutants in conserved sites (RP→AA, R→E, YFD→AAA, and Δ(KFHYFD)), and the interactions are not disrupted by mutation in the non-conserved sites (S→A at 136 a.a.). Moreover, the disruptions of the Set1-RNAPII interaction are at different degrees in mutants from two sites: the Set1 RP site mutants (RP→AA and R→E) still interact with RNAPII at a low level, while the interactions between RNAPII and Set1 KFHYFD site mutants (YFD→AAA and Δ(KFHYFD)) are below the detection of Western blot. Together, we termed Set1 amino acids 133-153 as Set1 RNA polymerase II interacting motif (SRIM), which is required for Set1-RNAPII interaction. However, this SRIM does not share homology with the SRI domain of Set2 or any other CTD binding proteins in *S. cerevisiae*.

Previously studies suggested that yeast Swd2 interacts with amino acids 1-229 of Set1 *in vitro*, and human Wdr82 (Swd2 homolog) is required for human Setd1A to interact with Ser5-P CTD peptides *in vitro* (115, 117). Thus, it is possible that the Set1 SRIM mutants could disrupt the interaction between Set1 and Swd2. To test this

possibility, vectors expressing Set1 and set1 SRIM mutants were transformed to *SWD2-3×HA, set1*Δ stain, and an immunoprecipitation assay was performed as described in Figure 2.3B. As shown in Figure 2.3C, Swd2 interacts with all the Set1 SRIM mutants, that disrupted the Set1-RNAPII interaction, suggesting that SRIM of Set1 is not required for Set1 interacting with Swd2.

To determine whether Set1 specifically interacts with the phosphorylation of Ser5 on CTD, not other domains of Rpb1 or other RNAPII subunits, we generated pRS415-*PYK1p*-NLS-GST-CTD fusion constructs, including GST tagged full-length wild-type CTD, GST tagged S2A CTD (all Ser2 mutated to Ala in 26 repeats), or GST tagged S5A CTD (all Ser5 mutated to Ala in 26 repeats). These vectors were transformed to wild-type, *SET2-3×FLAG*, and *3×FLAG-SET1* strains. The whole cell extracts of indicated strains were prepared and subjected to an α-FLAG (M2) immunoprecipitation assay and followed by Western blot using α-FLAG (M2), α-Ser5P (3E8), and α- Ser2P (H5) antibodies. As shown in Figure 2.4A, the blots for inputs show the presence of Ser5-P in wild-type CTD and S2A CTD, not in S5A CTD, and the result shows Set1 interacts with wild-type CTD and S2A CTD, not S5A CTD. To further determine whether set1 SRIM is required for this interaction, the pRS415-*PYK1p*-NLS-GST-CTD vector was transformed to the *set1*Δ strain with full length Set1 or set1 YFD→AAA mutant expressed from pRS416. As shown in Figure 2.4B, Set1 interacts with wild-type CTD while Set1 YFD→AAA mutant fails to. These results indicate that Set1 specifically interacts with the phosphorylation of Ser5 on CTD, and this interaction requires SRIM.

2.3.3 The SRIM is required for Set1 localization and H3K4 trimethylation levels at 5' of actively transcribed gene

Given the Set1 and RNAPII interaction is potentially important for Set1 function, we sought to investigate the consequences of Set1 SRIM mutants in yeast. To determine whether the set1 SRIM mutants cause a loss of global H3K4 methylation, the whole cell extracts of indicated strains were blotted with antibodies specifically against H3K4 mono-, di-, and trimethylation. With H3 levels as the loading controls, the results showed that set1 mutants have distinct global levels of H3K4 trimethylation: set1 YFD→AAA mutant has a great reduction in global H3K4 trimethylation level, whereas the R→E mutant has a slightly decrease or no change in global H3K4 trimethylation level comparing to wild-type *SET1* strain (Figure 2.5A). The difference may be due to set1 R→E mutant partially disrupt Set1-RNAPII interaction (Figure 2.3B and C) Interestingly, the global mono- and dimethylation levels of the set1 SRIM mutants are not changed compared to wild-type *SET1* strain (Figure 2.5A). The decrease of global H3K4 trimethylation is not due to loss of Set1 protein, for the Set1 protein levels of different mutants do not show difference with Set1 (Figure 2.5A).

To determine whether Set1 SRIM mutants disrupt Set1 localization at actively transcribed genes, an α-FLAG (M2) ChIP analysis was performed using the indicated stains. As shown in Figure 2.5C, the wild-type Set1 is predominantly enriched at 5' over promoters and 3' of *PYK1* gene compared to untagged wild-type background; whereas set1 SRIM (R→E and YFD→AAA) mutants are no longer significantly enriched at 5' over promoters and 3' of *PYK1* gene (Figure 2.5C). Our data indicates that the Set1-

RNAPII interaction is required for Set1 association with 5’ of active gene. Then we sought out to determine whether the set1 SRIM mutants affect H3K4 methylation deposition by ChIP assay using histone specific antibodies. As shown in Figure 2.5D, The H3K4 trimethylation distribution in set1 YFD→AAA mutant is overall greatly decreased on promoter, 5’ and 3’ of *PYK1* gene, however the H3K4 trimethylation in set1 R→E mutant shares a similar level with Set1. This is correlated with the global H3K4 trimethylation levels in Figure 2.5A. The difference may be also due to set1 R→E mutant partially disrupt Set1-RNAPII interaction (Figure 2.3B and C) As for H3K4 dimethylation, both set1 SRIM mutants shows no significant change compared to wild-type Set1. Besides *PYK1* gene, we also performed the ChIP analysis on another active gene, *PMA1*, which shows similar results as *PYK1* gene (Figure 2.5F, G, H, and I). Together, because set1 R→E mutant still partially interacts with RNAPII, we conclude that Set1-RNAPII is required for proper Set1 association with 5’ of active gene and the deposition of H3K4 trimethylation at 5’ of active gene, while the H3K4 dimethylation of active gene is not dependent by Set1-RNAPII interaction.

2.3.4 Set1 SRIM mutants disrupt Set1 recruitment and H3K4me3 at the 5’ of the induced *PCK1* gene.

To further investigate whether Set1 SRIM mutants affect Set1 recruitment and H3K4 methylation deposition, we looked in yeast gene at inducible condition. When the carbon souse of yeast medium is switched from glucose to acetate, the gluconeogenesis pathway will be turned on and the transcript levels of *Phosphoenolpyruvate carboxykinase* (*PCK1*) gene is quickly increased in both wild-type and *set1*Δ (Figure

2.6A). The Set1 protein levels and global H3K4 trimethylation remain unchanged under acetate condition for tested 4 hours (Figure 2.6B). The 3×FLAG-Set1 is increasingly recruited to the promoter and 5' of *PCK1* gene from 0 to 2 hours (Figure 2.6D) And H3K4 trimethylation is increasingly enriched at promoter and 5' of *PCK1* gene from 0 to 4 hours (Figure 2.6F).

To determine whether the *PCK1* transcription will be affected by set1 SRIM mutants, the transcript levels of *PCK1* in set1 mutants under inducible condition were determined by qRT-PCR (Figure 2.7A). The statistical analysis shows there is no significant difference between set1 mutants with Set1 when *PCK1* is turned on for 4 hours. In addition, the Set1 protein levels in Set1 and set1 mutants remain unchanged for tested 2 hours (Figure 2.7B). And the global levels of H3K4 mono-, di- and trimethylation in different mutants are not changed from their original levels during the medium switch for 2 hours (Figure 2.7B).

Interestingly, as shown in Figure 2.8A and C and compared to full-length Set1, set1 SRIM mutants (R→E and YFD→AAA) fail to be recruited at promoter and 5' of *PCK1* gene and in these two mutants the H3K4 trimethylation are failed to be enriched at *PCK1* gene when *PCK1* gene is turned on. The *PCK1* is turned on normally as there is no defect of Rpb3 recruitment on *PCK1* gene (Figure 2.7A and 2.8B). This suggests the role of RNAPII in recruiting Set1 to deposit H3K4 trimethylation at the gene when the gene is turned on. However, the function of the H3K4 trimethylation is still unknown.

As shown in Figure 2.8D, set1 YFD→AAA mutant fails to establish H3K4me2 and set1 R→E mutant has a delay to establish H3K4me2 at promoter and 5' ends of

PCK1 gene for tested 2 hours when *PCK1* gene is turned on. This suggests the role of RNAPII in recruiting Set1 to deposit H3K4 dimethylation at the gene when the gene is turned on.

2.3.5 Set1 SRIM mutants have growth defect on 6-AU

6-azouracil (6-AU) is a drug that decreases the intracellular GTP and UTP pools, resulting in yeast growth sensitivity when there is a defect in transcriptional elongation. Based on studies that the deletion of *SET1* or H3K4R mutant showed growth defects on 6-azouracil (6-AU) and Set1 SRIM mutants affect Set1 recruitment and H3K4 trimethylation deposition, we sought to determine whether Set1 SRIM mutants have growth defects on 6-AU (111). As shown in Figure 2.9A, *set1*Δ with vector has a slow growth defect on 150 μg/ml 6-AU plates at 30°C, and this defect is more severe when the plate is placed at 37°C. The full length Set1 expressed from *CEN* vector can rescue this growth defect at both temperatures, while Set1 H1017K catalytic mutant cannot. However, the Set1 R→E and YFD→AAA mutant still have growth defects on 6-AU, more severely at 37°C. Interestingly, the Set1 R→E mutant grew a better than YFD→AAA mutant, which is correlated with the global H3K4 trimethylation level of Set1 R→E mutant is higher than YFD→AAA mutant.

2.3.6 Overexpression of set1$_{1\text{-}375}$ exhibits slow growth phenotype

As shown in Figure 2.2A, set1$_{1\text{-}375}$ is necessary and sufficient to interact with RNAPII. It is possible that overexpression this domain results in a gain of function to compete with endogenous Set1 binding to RNAPII. Thus, we generated pRS425-*PYK1p-set1*(1-375) and pRS425-*PYK1p-set1*(1-375) YFD→AAA mutant and transformed them

into wild-type and *set1*Δ strain. Overexpression of set1$_{1-375}$ exhibits a slightly slow growth phenotype in both wild-type and *set1*Δ strain, and overexpression of set1$_{1-375}$ YFD→AAA mutant does not show this slow growth phenotype. This indicates the slow growth phenotype is independent of endogenous Set1 and it is possible that the slow growth phenotype is through set1$_{1-375}$ competing with other important proteins for Ser5-P CTD of RNAPII.

2.4 Discussion

2.4.1 Identification and characterization of how Set1 interacts with RNAPII containing Ser5-P CTD.

In this study, we determined that Set1 interacts Ser5-P and/or Ser7-P CTD of RNAPII, whereas Set2 binds to Ser2-P, Ser5-P, and/or Ser7-P CTD of RNAPII. It suggests that besides Ser2-P and Ser5-P, the phosphorylation on Ser7 in CTD also plays a role in recruiting histone methyltransferases. In addition, we determined that Set1-RNAPII interaction is independent of Paf1 complex and most Set1-associated subunits. This contradicts with the current model that RNAPII recruitment of Set1 is mediated by Paf1 complex (150-152). However, in these previous studies, it is determined the dependency of paf1 complex in Bre2-RNAPII interaction (150, 151). As shown in Figure 2.1C, Set1-RNAPII interaction is independent of Bre2. Our data is supported by the study that Setd1A, Set1 homolog in human, fails to be co-immunoprecipitated with Paf1 (117).

We also determined the SRIM of Set1 and requirement of SRIM for Set1-RNAPII interaction. The unique feature of this motif is that SRIM mutants only disrupt Set1-

RNAPII interaction without disrupting Set1-Swd2 interaction. The study of interactions between Set1 and Set1-associated subunits shows that only Swd2 binds to N terminus of Set1 (amino acids 1-229) and all the other Set1 complex subunits associate with downstream of that region (115). Thus, the SRIM mutants provide valuable tools to study the Set1 function independent of RNAPII recruitment with all Set1-assoicated subunits on board.

By protein sequence comparison, the human RRM domain in Setd1A exhibits 23% identity and 42% similarity to yeast Set1 RRM1 domain, however, the SRIM cannot be identified in Setd1A, suggesting the Set1 proteins are quite divergently evolved. The human Setd1A RRM and Ser5-P CTD interaction is dependent of Wdr82 *in vitro* (117). In our study, we did not rule out the possibility that Swd2 is required for Set1-RNAPII interaction, however, our data shows that without Set1, Swd2 does not bind to RNAPII. Because Swd2 also functions in the cleavage and polyadenylation factor complex which is essential for yeast, and the deletion of *SWD2* can be achieved by overexpression C-terminus of Sen1 (164). To determine whether Swd2 is required for Set1-RNAPII interaction, immunoprecipitation assay of Set1 with RNAII in deletion strain of *SWD2* and overexpression C-terminus of Sen1 and *in vitro* binding assay of Set1, Swd2, and Ser5-P CTD are needed in future.

2.4.2 Characterization the dependency of RNAII in Set1 recruitment and Histone H3K4 methylation deposition

Our set1 SRIM mutants show a great reduction of Set1-chromatin association at 5’ without affecting set1 mutant localization at promoter and 3’ of actively transcribed

genes (Figure 2.5C), and set1 SRIM mutants fail to be recruited by RNAPII at induced *PCK1* gene. This is, for the first time, the direct evidence to support the model that Set1 is recruited by RNAPII containing Ser5-P CTD at 5' of active genes. However, the previous model is not complete: without RNAPII recruitment, set1 SRIM mutants are still enriched and evenly distributed at promoter, 5', and 3' of actively transcribed genes over the background, possibly revealing an unknown mechanism of Set1 recruitment to chromatin independent of RNAPII.

Our data also suggest that Set1 can be recruited to chromatin to deposit H3K4 trimethylation at transcription start site through a RNAII recruitment manner, as set1 YFD→AAA mutant has a great decrease in H3K4 trimethylation levels globally and gene specifically (Figure 2.10). Interestingly, the set1 R→E mutant, with partial disruption in Set1-chromatin association, has full levels of H3K4 trimethylation globally and at active gene. This could be explained by the R→E mutant still interacts with RNAPII containing Ser5-P at low level. It is possible that this weak interaction is enough to establish full trimethylation at actively transcribed genes in steady state and this transient and weak set1 mutant and chromatin interaction cannot be detected by ChIP.

Our combined data suggest that Set1 can also be recruited to chromatin and methylate H3K4 through other mechanisms independent of RNAII recruitment (Figure 2.10). This idea can be support by multiple studies from us and other labs that the C-terminus of Set1, containing nSET domain, SET domain, and post-SET domains, has full levels of mono- and dimethylation, and partial trimethylation activity *in vivo* (113, 115, 116). As shown in Figure 2.2B, C-terminus of Set1, containing amino acids 690-1080,

does not interacts with RNAPII. The RNAPII-independent Set1 recruitment could be through Set1-asscoiated subunits or other unknown mechanisms. Our previous data showed, set2 without SRI domain still binds to nucleosome and catalyze H3K36 monomethylation (165).

Interestingly, in set1 YFD→AAA mutant, H3K4 trimethylation is reduced overall across the ORF, however, it is still more enriched at 5' comparing to promoter and 3' ends, while Set1 is enriched evenly at promoter, 5' and 3' of ORF. The enrichment of H3K4 trimethylation at 5' of ORF in the YFD→AAA mutant is independent of Set1 occupancy and interaction between Set1 and RNAPII. The mechanism for this is still unknown. One possibility is that Set1 catalytic activity is activated at 5' due to stimulation by other factors. Another possibility is that the turnover speeds of H3K4 trimethylation marker are faster at promoter and 3' than 5' ends of ORF. By disrupting Set1-RNAPII interaction results in decrease of overall H3K4 trimethylation but H3K4 trimethylation is still enriched at 5' ends of ORF.

The 6-AU assay suggested there are potential transcriptional defects in set1 SRIM mutants. Further experiments would be needed to show transcriptional elongation is reduced, for example nuclear run-on assay.

2.5 Materials and Methods

Yeast strains and strain construction

All yeast strains were derived from *S. cerevisiae* BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) or BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*). Genotypes of all strains used in this study are described in Table 2.2. The *3×FLAG-SET1* was generated by using pK3F as template in BY4741. C-terminal 3×FLAG tagged Bre2, Swd1, Spp1, Swd2, Swd3, Sdc1, and Shg1 strains using p3FLAG vector (166). The *bre2*Δ*, swd1*Δ*, swd3*Δ, *spp1*Δ, *sdc1*Δ, *shg1*Δ, *paf1*Δ, *ctr9*Δ, *cdc73*Δ*, rtf1*Δ, and *leo1*Δ in *3×FLAG-SET1* were generated by PCR mediated gene amplified the *KanMX* cassette from Saccharomyces genome deletion project (167). A deletion allele of *SET1* is generated by PCR amplification of *HphMX* cassette from pAG32 (168).

Preparation of yeast whole-cell extracts, Western blot analysis and antibodies

Yeast extracts and Western blot analysis were performed as described previously (113, 119, 169). The α-FLAG (M2) mouse antibody (Sigma, F1804) was used at 1:5,000. The α-FLAG rabbit antibody (Sigma, F7425) was used at 1:5,000. The glucose-6-phoshate dehydrogenase rabbit antibody (Sigma, A9521) was used at 1:50,000. The α-H3K36me3 rabbit antibody (Abcam, Ab9050) was used at 1:5,000. The α-H3 rabbit antibody (Active Motif, 39163) was used at 1:5,000. The α-H3K4me3 rabbit antibody (Active Motif, 39159) was used at 1:100,000. 4E12 antibody (specific to P-Ser7 CTD, Emdmillipore, 04-1570) was used 1:5,000, H14 antibody (specific to Ser5P CTD, Covance, MMS-134R) was used at 1: 10,000. H5 antibody (specific to P-Ser2 CTD, Covance, MMS-129R) were used at 1:2,000 for IP and 1:10,000 for input. 3E8 (α-Ser5P, Emdmillipore, 04-1572) was used 1:5,000.

Gene expression analysis

RNA transcript levels were determined using quantitative real-time PCR (qRT-PCR) as previously described (133). Three technical repeats were performed for each of the three biological repeats. The All ΔΔCT values were analyzed by normalizing values to *ACT1* or *MDH3* and transcript levels of *SET1* were compared relative to the transcript level in a WT strain. The primer sequences used can be found in Table 2.1.

Immunoprecipitation assay

The overnight cultures of indicated yeast cells were diluted to an OD600 of 0.01 and regrew up to an OD600 of 0.6-0.8, 100 mL cultures were harvested, and quickly followed by whole-cell extracts by beads beating in by using lysis Buffer (20mM HEPES pH8.0, 350mM NaCl, 10% glycerol, 0.1% Tween 20, with freshly added 1µl/ml pepstatin (1 mg/ml), 1 µl/ml Aprotinin/ leupeptin (1 mg/ml), 1 µl/ml PMSF (100mM), 20mM Sodium fluoride, and 20mM sodium orthovanadate). 2 µl α-FLAG (M2) monoclonal antibody (Sigma F1804) was conjugated to 10 µl Protein-G magnetic beads (Dynabeads, Life Technologies 10004D) for about 30 min. The extracts were incubated with pre-equilibrated α-FLAG (M2) conjugated magnetic beads for 2 h at 4°C, then the beads were washed three times with 1ml of lysis Buffer containing protease and phosphatase inhibitors mentioned above for 5 min each. The beads were boiled in SDS-PAGE loading buffer, and bead-bound proteins were separated by a magnet and analyzed by immunoblotting with antibodies directed against the phosphorylated CTD and α-FLAG (M2) antibody.

ChIP assay

For normal glucose cell grow condition, the condition is same as immunoprecipitation assay. For acetate inducible condition, all strains were grown overnight to saturation in SC-Ura media. They were then back diluted to OD600 0.01 in 100 ml SC-Ura media and grown at 30°C to mid-log phase (OD600 0.6), the cells were suspended and washed once with SC-Ura with 1% potassium acetate (no glucose). Then the cells were regrown in 100 ml SC-Ura with 1% potassium acetate for indicated time. The ChIP assay procedure was ZipChIP method as described previously (170). The primer sequences are in Table 2.1.

6AU growth assay

Indicated yeast strains were grown in synthetic complete drop-out media lacking uracil (SC-Ura). Overnight cultures were diluted to an OD600 of 0.1 and regrown to an OD600 of 0.6, and fivefold serial dilutions were plated on SC-Ura plates with or without 6-AU (Sigma) at final concentration of 150 µg/ml. Cells were grown at 30°C and 37°C before plates were photographed.

Table 2.1 Primer Sequences in Chapter 2

Name	Sequence
PYK1 promoter (ChIP)	/56-FAM/CGAATATCG/ZEN/TTTTGATGGCGAGCCTTT/3IABkFQ/ AGAGGTTCTTGGAAATGAAAAGTTAC CCTTCCCATATGATGCTAGGTAC
PYK1 5' (ChIP)	/56-FAM/AGAATTGTA/ZEN/CCCAGGTAGACCATTGGC/3IABkFQ/ TTCTGATTTCTGGACCCTTGG CATTGACAACGCCAGAAGTC
PYK1 3' (ChIP)	/56-FAM/CTCTCACTT/ZEN/GTACAGAGGTGTCTTGGGA/3IABkFQ/ TCAGAGACAGGTTCCTTTTCG TTACCAGATGCCCAAGAGC
PCK1 (-360bp) (ChIP)	/56-FAM/CCCAAACAG/ZEN/GATTGTAAAGCTTAGACGC/3IABkFQ ACGTACCATTGTCCAACCAG ATACAAGGGAAAGATGCCGG
PCK1 (-197bp) (ChIP)	/56-FAM/AGGGCAGAG/ZEN/AAGTCTTTGTAATGTGTGT /3IABkFQ/ ACTGATTACACATGGAAGAGAAGT TAGCAACATTGGGCAGAGTATAG
PCK1 5' (120bp) (ChIP)	/56-FAM/TGCTCCAGC/ZEN/TGCCGTTTTGTATGA/3IABkFQ/ ATTGGCTCTTAGTGACGAAGTC CACCGGAATAAGCGATCAATG
PCK1-F (Gene expression)	TCAGACAAGAATTGGCTCTTAGTGA
PCK1-R (Gene expression)	ACGGCAGCTGGAGCATTG
MDH3-F (Gene expression)	GAGCTGGCGTTGTACGATATCC
MDH3-R (Gene expression)	TTGGTGTTGATGTGAGATAAATCCTT

Table 2.2 Yeast Strains and Genotype in Chapter 2

Yeast Strain	Genotype	Reference
BY4741 (wild-type)	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0*	Open Biosystems
Set2-*3×FLAG*	BY4742 *SET2-3×FLAG::KanMX*	(141)
Set2-(*ΔSRI*)*3×FLAG*	BY4742 *SET2$_{1-618}$-3×FLAG::KanMX*	(141)
SDBY1420	BY4741 *set1Δ::HphMX*	This study
SDBY1421	BY4741 *SET1::loxP-3×FLAG-SET1*	This study
SDBY1501	BY4741 *BRE2::BRE2-3×FLAG*	This study
SDBY1502	BY4741 *BRE2::BRE2-3×FLAG*, *set1Δ::HphMX*	This study
SDBY1503	BY4741 *SWD1::SWD1-3×FLAG*	This study
SDBY1504	BY4741 *S* *set1Δ::HphMX*	This study
SDBY1505	BY4741 *SPP1:: SPP1-3×FLAG*	This study
SDBY1506	BY4741 *SPP1:: SPP1-3×FLAG*, *set1Δ::HphMX*	This study
SDBY1507	BY4741 *SWD2:: SWD2-3×FLAG*	This study
SDBY1508	BY4741 *SWD2:: SWD2-3×FLAG*, *set1Δ::HphMX*	This study
SDBY1509	BY4741 *SWD3:: SWD3-3×FLAG*	This study
SDBY1510	BY4741 *SWD3:: SWD3-3×FLAG*, *set1Δ::HphMX*	This study
SDBY1511	BY4741 *SDC1:: SDC1-3×FLAG*	This study
SDBY1512	BY4741 *SDC1:: SDC1-3×FLAG*, *set1Δ::HphMX*	This study
SDBY1513	BY4741 *SHG1:: SHG1-3×FLAG*	This study
SDBY1514	BY4741 *SHG1:: SHG1-3×FLAG*, *set1Δ::HphMX*	This study
SDBY1515	BY4741 *SET1::loxP-3×FLAG-SET1*, *bre2Δ::KanMX*	This study
SDBY1516	BY4741 *SET1::loxP-3×FLAG-SET1*, *swd1Δ::KanMX*	This study
SDBY1517	BY4741 *SET1::loxP-3×FLAG-SET1*, *swd3Δ::KanMX*	This study
SDBY1518	BY4741 *SET1::loxP-3×FLAG-SET1*, *spp1Δ::KanMX*	This study
SDBY1519	BY4741 *SET1::loxP-3×FLAG-SET1*, *sdc1Δ::KanMX*	This study
SDBY1520	BY4741 *SET1::loxP-3×FLAG-SET1*, *shg1Δ::KanMX*	This study

Table 2.2 continued

Table 2.2 continued

SDBY1521	BY4741 *SET1::loxP-3×FLAG-SET1*, *paf1Δ::KanMX*	This study
SDBY1522	BY4741 *SET1::loxP-3×FLAG-SET1*, *ctr9Δ::KanMX*	This study
SDBY1523	BY4741 *SET1::loxP-3×FLAG-SET1*, *rtf1Δ::KanMX*	This study
SDBY1524	BY4741 *SET1::loxP-3×FLAG-SET1*, *leo1Δ::KanMX*	This study
SDBY1525	BY4741 *RPB3::RPB3-3HA, set1Δ::HphMX*	This study
SDBY1526	BY4741 *SWD2::SWD2-3HA, set1Δ::HphMX*	This study

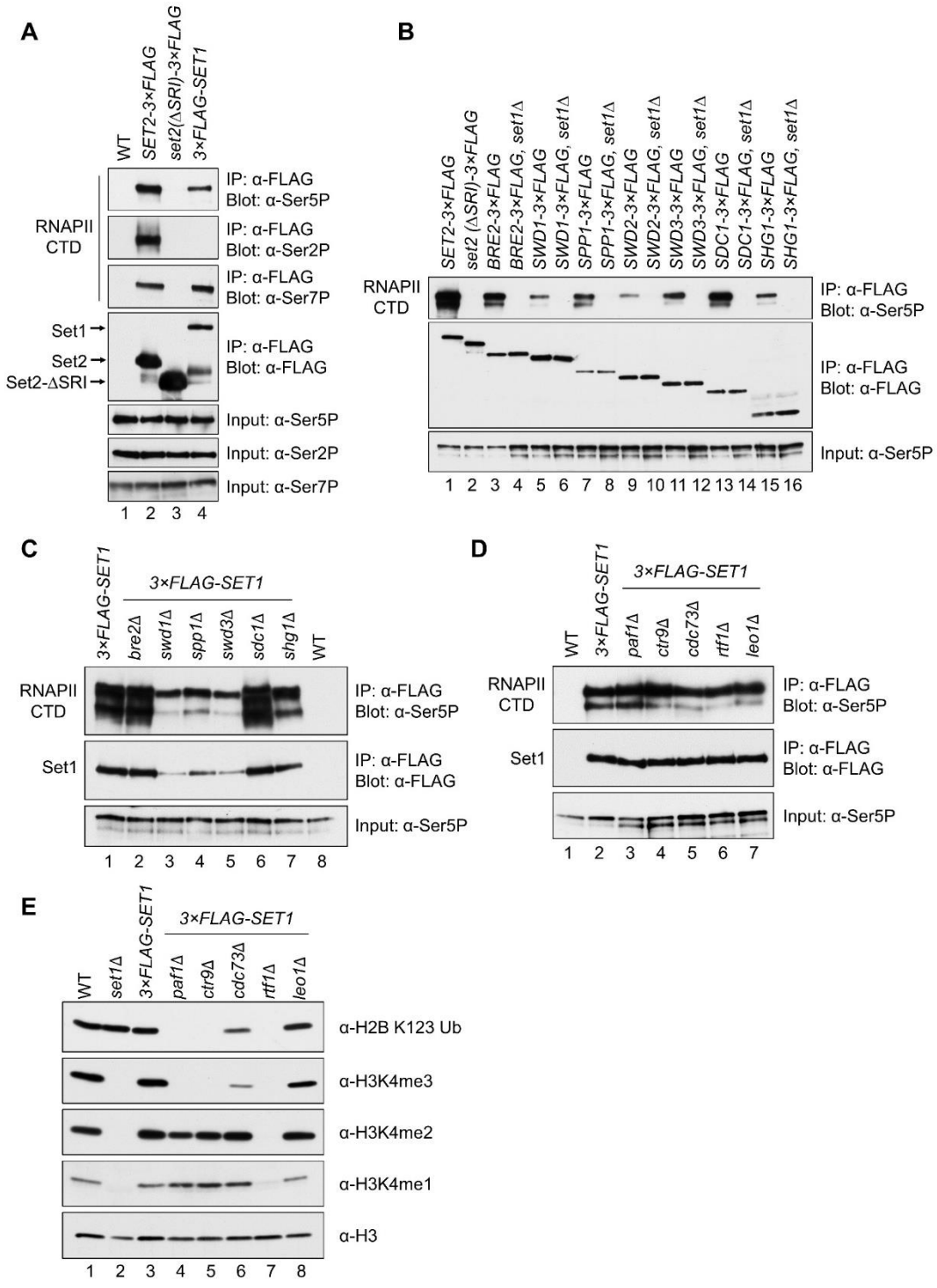


Figure 2.1 Set1 interacts with RNAPII containing Ser5-P and Ser7-P CTD, and this interaction is independent of most Set1 complex subunits and Paf1 complex subunits.

(A) Immunoprecipitation assay shows that Set1 interacts with RNAPII containing Ser5-P and/or Ser7-P CTD. FLAG tagged Set1, Set2, and set2(ΔSRI) are expressed from their endogenous promoters. Whole cell extracts were prepared and subjected to α-FLAG (M2) immunoprecipitation and followed by Western blotting using antibodies specific to Serine phosphorylated CTD (α-Ser5P (H14), α-Ser2P (H5), and α-Ser7P (4E12)) and M2 α-FLAG antibody. Set2-3×FLAG was used as a positive control, and WT and set2(ΔSRI)-3×FLAG were used as negative controls. The whole cell lysates were blotted with α-Ser5P, α-Ser2P and α-Ser7P for input levels. (B) Immunoprecipitation assay shows that Set1 is required for the interactions between Set1 complex subunits and RNAPII. C-terminal 3×FLAG tagged *BRE2*, *SWD1*, *SPP1*, *SWD2*, *SWD3*, *SDC1*, and *SHG1* strains were generated in WT and *set1*Δ strains. Whole cell extracts were prepared and subjected to α-FLAG (M2) immunoprecipitation and followed by Western blot using α-Ser5P (H14) and rabbit α-FLAG antibodies. Set2-3×FLAG was used as a positive control, and set2(ΔSRI)-3×FLAG were used as negative control. The whole cell lysates were blotted with α-Ser5P for input levels. The double bands of CTD of RNAPII is due to different phosphorylation levels of RNAPII. (C) Immunoprecipitation assay shows that Set1-RNAPII interaction is independent of most Set1 complex subunits. The Set1 complex subunits deletion strains *bre2*Δ, *swd1*Δ, *swd3*Δ, *spp1*Δ, *sdc1*Δ, and *shg1*Δ were generated in *3×FLAG-SET1* strain. Whole cell extracts were prepared and subjected to α-FLAG (M2) immunoprecipitation and followed by Western blotting using α-Ser5P (H14) M2 α-FLAG antibodies. WT was used as a negative control. The whole cell lysates were blotted with α-Ser5P for input levels. (D) Immunoprecipitation assay shows that Set1-RNAPII interaction is independent of Paf1 complex. The individual deletion of Paf1 complex subunits *paf1*Δ, *ctr9*Δ, *cdc73*Δ, *rtf1*Δ, and *leo1*Δ were generated in *3×FLAG-SET1* strain. Whole cell extracts were prepared and subjected to α-FLAG (M2) immunoprecipitation and followed by Western blotting using α-Ser5P (H14) M2 α-FLAG antibodies. WT was used as a negative control. The whole cell lysates were blotted with α-Ser5P for input levels. (E) Deletions of individual Paf1 complex subunits affect global H2B monoubiquitination and H3K4 methylation. Yeast whole cell extracts of indicated strains were blotted for H2B K123ub and methyl specific antibodies for H3K4 methylation status; and α-H3 for loading control.

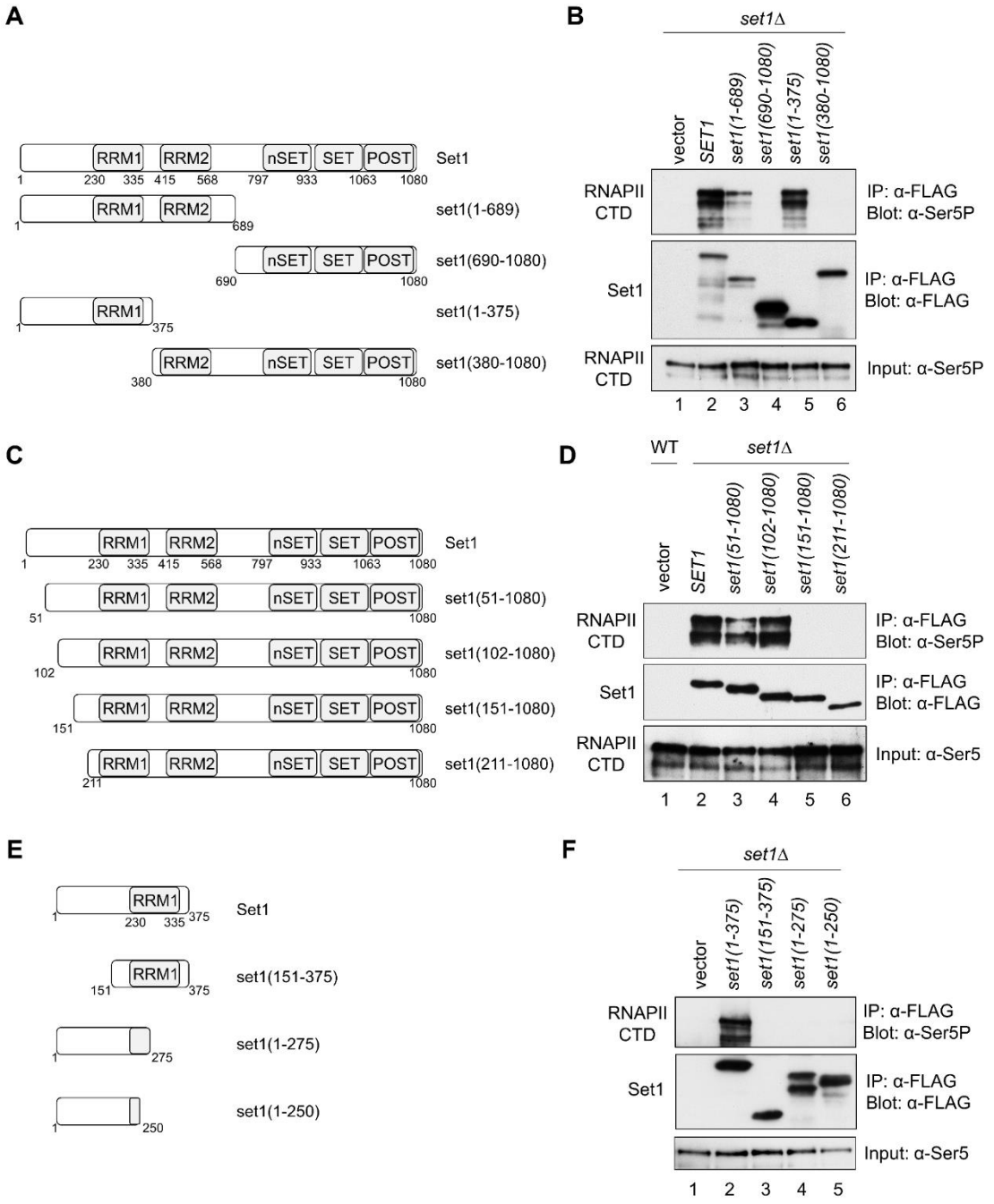


Figure 2.2 Identification of a novel region in Set1 required for RNAPII interaction.

(A) Schematic representation of the Set1 constructs used to determine the Set1-RNAPII interacting region. The domain structure of Set1 contains two RNA recognition motifs (RRM1 and RRM2), an n-SET domain, a SET domain, and a post-SET domain. (B) Immunoprecipitation assay shows that $Set1_{1\text{-}375}$ is necessary and

sufficient to interact with RNAPII containing Ser5-P CTD. 3×FLAG tagged Set1 and indicated set1 truncations were expressed from *CEN* plasmids under *SET1* promoter. The whole cell extracts were prepared and subjected to α-FLAG (M2) immunoprecipitation assay and followed by Western blotting using α-Ser5P(H14) and rabbit α-FLAG antibodies. The *set1*Δ with vector was used as negative control and the whole cell lysates were blotted with α-Ser5P for input levels. (C) Schematic representation of Set1 N-terminal sequential truncation constructs used to further determine the Set1-RNAPII interacting region. (D) Immunoprecipitation assay shows that Set1 amino acids 102-150 are required for interacting with RNAPII containing Ser5-P CTD. Set1 and set1 truncations are expressed under *SET1* promoter from *CEN* plasmids. The immunoprecipitation assay was performed as described in Figure 2.2B. WT with vector was used as negative control and the whole cell lysates were blotted with α-Ser5P for input levels. (E) Schematic representation of Set1 truncation constructs used to further determine the Set1-RNAPII interacting region. (F) Immunoprecipitation assay shows that both RRM1 and the uncharacterized region upstream of RRM1 are required for Set1-RNAPII interaction. 3×FLAG tagged set1 truncations were expressed under *PYK1* promoter from 2 micron plasmids. The immunoprecipitation assay was performed as described in Figure 2.2B. The *set1*Δ with vector was used as negative control and the whole cell lysates were blotted with α-Ser5P for input levels.

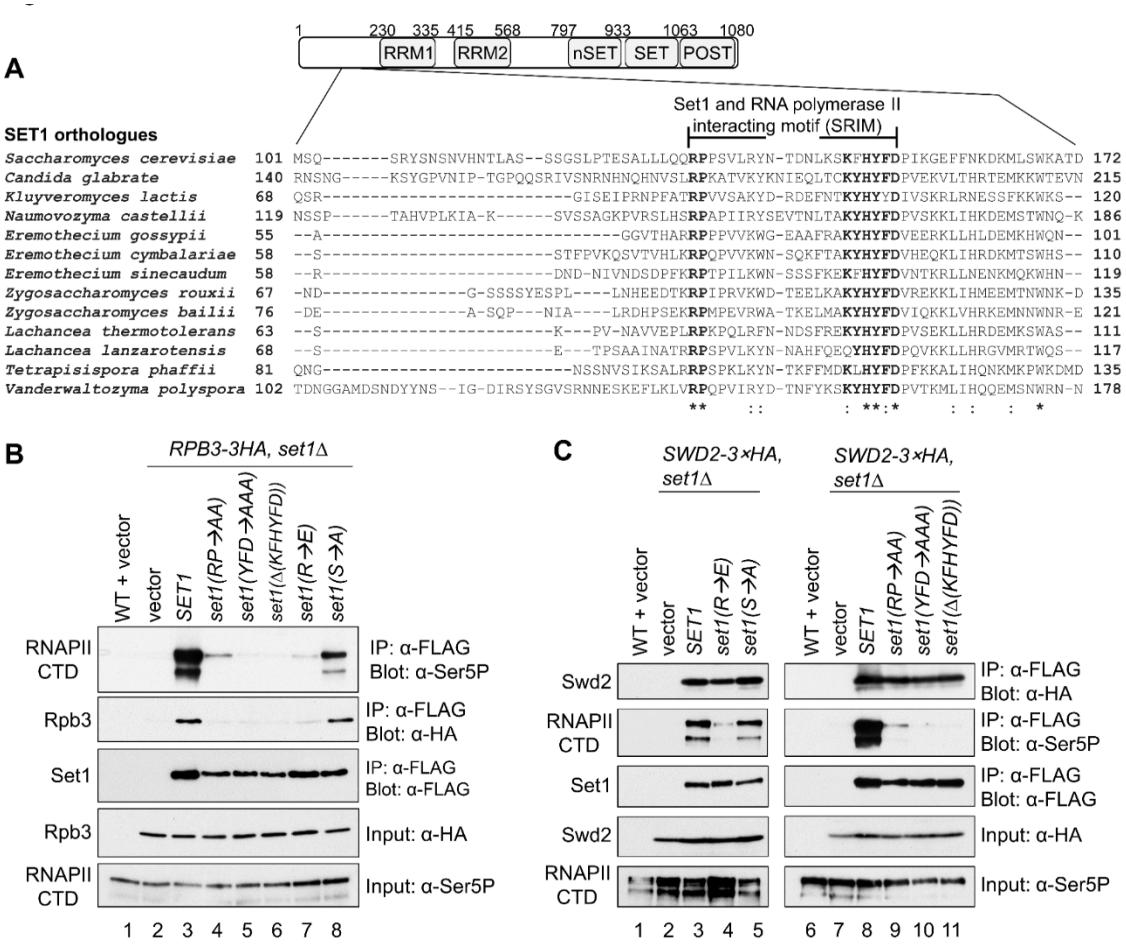


Figure 2.3 Conserved Set1 N-terminal residues are required for the Set1 interaction with the Ser5-P CTD of RNAPII.

(A) Sequence alignment of N-terminal uncharacterized regions of Set1 and its homologs from *Saccharomycetaceae* family. Sequences were aligned using Clustal Omega. An asterisk (*) indicates the position has a fully conserved residue; and a colon (:) indicates the position has high similarity between different groups. The most conserved residues in SRIM were bold. (B) Co-immunoprecipitation assay shows Set1 mutants in RP and KFHYFD sites disrupt Set1-RNAPII interaction. The constructs containing Set1 or indicated set1 mutant were transformed to *RPB3-3×HA, set1*Δ strain. The whole cell extracts were prepared and subjected to α-FLAG (M2) immunoprecipitation assay and followed by Western blot using antibodies α-HA (12CA5) for Rpb3, α-Ser5P (H14) for CTD of RNAPII, M2 α-FLAG for Set1. WT with vector was used as a negative control and the whole cell lysates were blotted with α-Ser5P and α-HA for input levels. (C) Co-immunoprecipitation assay shows set1 SRIM mutants still interact with Swd2. The constructs containing Set1 or indicated set1 mutant were transformed to *SWD2-3×HA, set1*Δ strain. The immunoprecipitation assay was performed as described in Figure 2.3B using antibodies α-HA (12CA5) for Swd2, α-Ser5P (H14) for CTD of RNAPII, M2 α-FLAG for Set1. WT with vector was

used as a negative control and the whole cell lysates were blotted with α-Ser5P and α-HA for input levels.

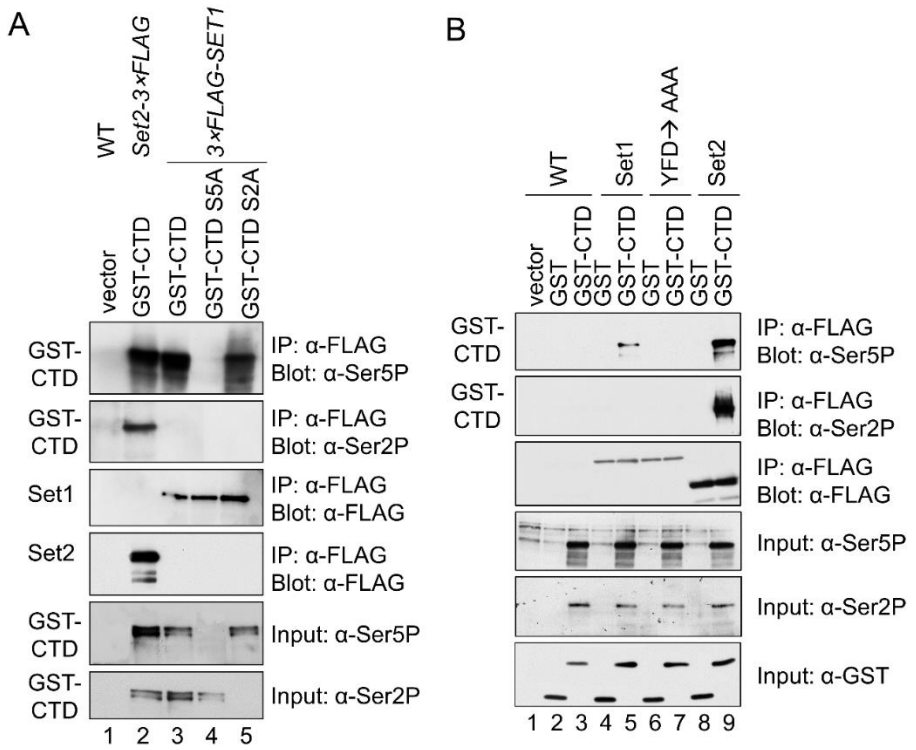


Figure 2.4 Set1 interacts with the phosphorylation of Ser5 on CTD repeats *in vivo*.

(A) Co-immunoprecipitation assay shows Set1 specifically interacts with the phosphorylation of Ser5 on CTD repeats. The *PYK1p*-NLS-GST-CTD fusion constructs containing wild-type, S2A, or S5A CTD repeats were transformed to indicated strains. The whole cell extracts were prepared and subjected to α-FLAG (M2) immunoprecipitation assay and followed by Western blot using α-Ser5P(3E8), α-Ser2P(H5), and M2 α-FLAG antibodies. The wild-type with vector was used as a negative control and the whole cell lysates were blotted with α-Ser5P and α-Ser2P for input levels. (B) Co-immunoprecipitation assay shows set1 SRIM mutant disrupts Set1 interacting with the phosphorylation of Ser5 on CTD repeats. FLAG immunoprecipitations were blotted using α-Ser5P, α-Ser2P, and M2 FLAG. The WT with vector was used as a negative control and the whole cell lysates were blotted with α-Ser5P, α-Ser2P, and α-GST for input levels. The immunoprecipitation assay of indicated strains was performed as described in Figure 2.4A. The wild-type with vector was used as a negative control and the whole cell lysates were blotted with α-Ser5P, α-Ser2P, and α-GST for input levels.

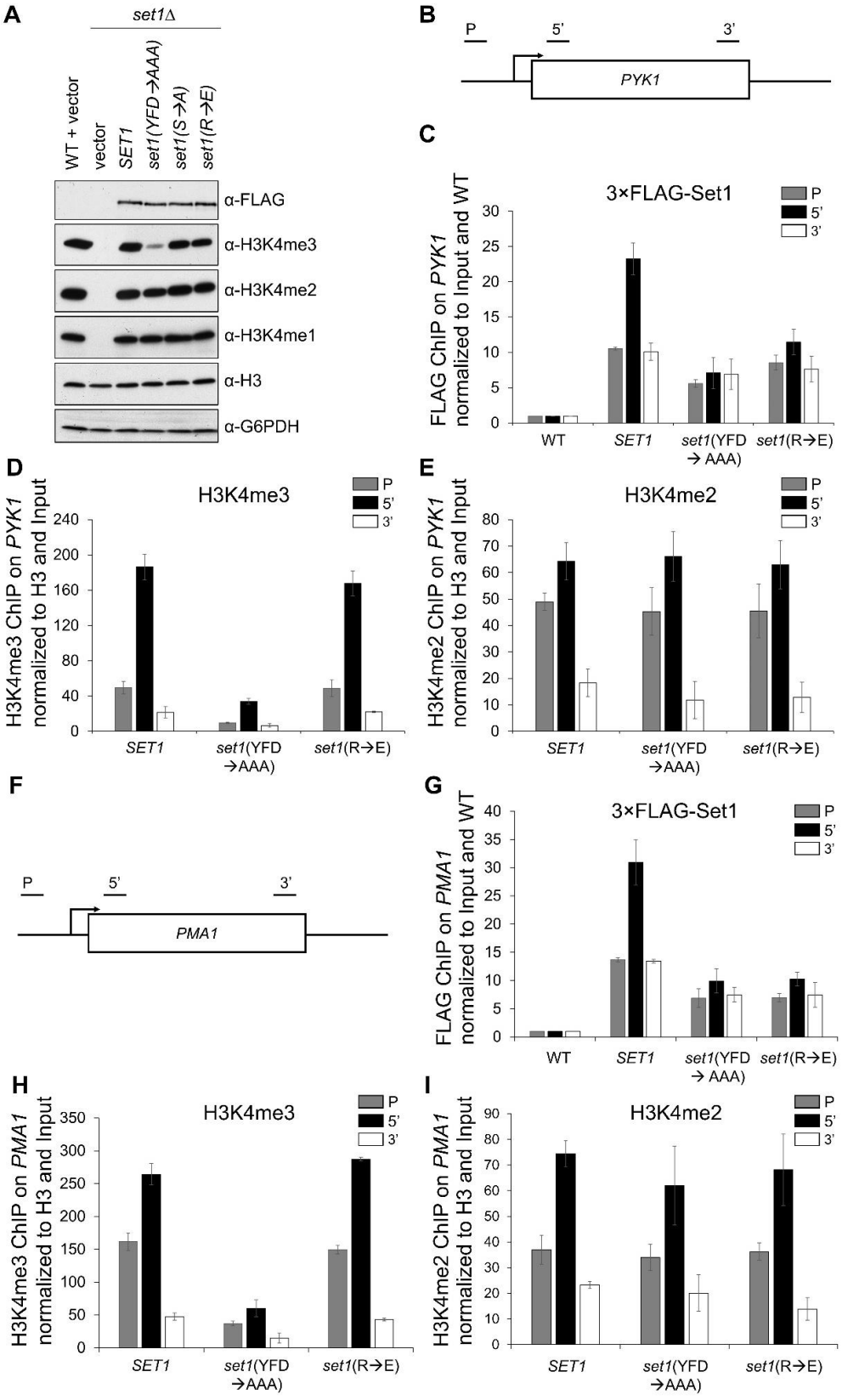

A
set1Δ
WT + vector
vector
SET1
set1(YFD→AAA)
set1(S→A)
set1(R→E)
α-FLAG
α-H3K4me3
α-H3K4me2
α-H3K4me1
α-H3
α-G6PDH
B
P
5'
3'
PYK1
C
3×FLAG-Set1
FLAG ChIP on PYK1 normalized to Input and WT
WT
SET1
set1(YFD → AAA)
set1(R→E)
D
H3K4me3
H3K4me3 ChIP on PYK1 normalized to H3 and Input
E
H3K4me2
H3K4me2 ChIP on PYK1 normalized to H3 and Input
F
PMA1
G
3×FLAG-Set1
FLAG ChIP on PMA1 normalized to Input and WT
H
H3K4me3
H3K4me3 ChIP on PMA1 normalized to H3 and Input
I
H3K4me2
H3K4me2 ChIP on PMA1 normalized to H3 and Input

Figure 2.5 The SRIM is required for Set1 localization and H3K4 trimethylation levels at 5' of actively transcribed gene

(A) Set1 protein levels are not changed in set1 SRIM mutants. Yeast whole cell extracts of indicated strains were blotted for α-FLAG (M2) for Set1 protein level, and methyl specific antibodies for H3K4 methylation status, α-H3 and α-G6PDH for loading control. (B) Schematic of regions analyzed for ChIP using specific primers across the *PYK1* gene. (C) ChIP analysis of 3×FLAG tagged Set1 and set1 SRIM mutants was performed using M2 α-FLAG antibody, and the signal is normalized to input and relative to WT. (D) ChIP analysis of *3×FLAG-SET1* and *set1* SRIM mutants was performed using α-H3K4me3 antibodies, respectively, and the signal is normalized to H3 and input levels, then relative to the signal of *set1*Δ strain. Three technical repeats were performed for each of the three biological repeats in all ChIP analysis in Figure 2.5 and the error bars represent the standard error of the mean. (E) ChIP analysis of *3×FLAG-SET1* and *set1* SRIM mutants was performed using α-H3K4me2 antibodies, respectively, and the signal is normalized to H3 and input levels, then relative to the signal of *set1*Δ strain. Three technical repeats were performed for each of the three biological repeats in all ChIP analysis in Figure 2.5 and the error bars represent the standard error of the mean. (F) Schematic of regions analyzed for ChIP using specific primers across the *PMA1* gene. (G) ChIP analysis of 3×FLAG tagged Set1 and set1 SRIM mutants was performed using M2 α-FLAG antibody, and the signal is normalized to input and relative to WT. (H) ChIP analysis of *3×FLAG-SET1* and *set1* SRIM mutants was performed using α-H3K4me3 antibodies, respectively, and the signal is normalized to H3 and input levels, then relative to the signal of *set1*Δ strain. Three technical repeats were performed for each of the three biological repeats in all ChIP analysis in Figure 2.5 and the error bars represent the standard error of the mean. (I) ChIP analysis of *3×FLAG-SET1* and *set1* SRIM mutants was performed using α-H3K4me2 antibodies, respectively, and the signal is normalized to H3 and input levels, then relative to the signal of *set1*Δ strain. Three technical repeats were performed for each of the three biological repeats in all ChIP analysis in Figure 2.5 and the error bars represent the standard error of the mean.

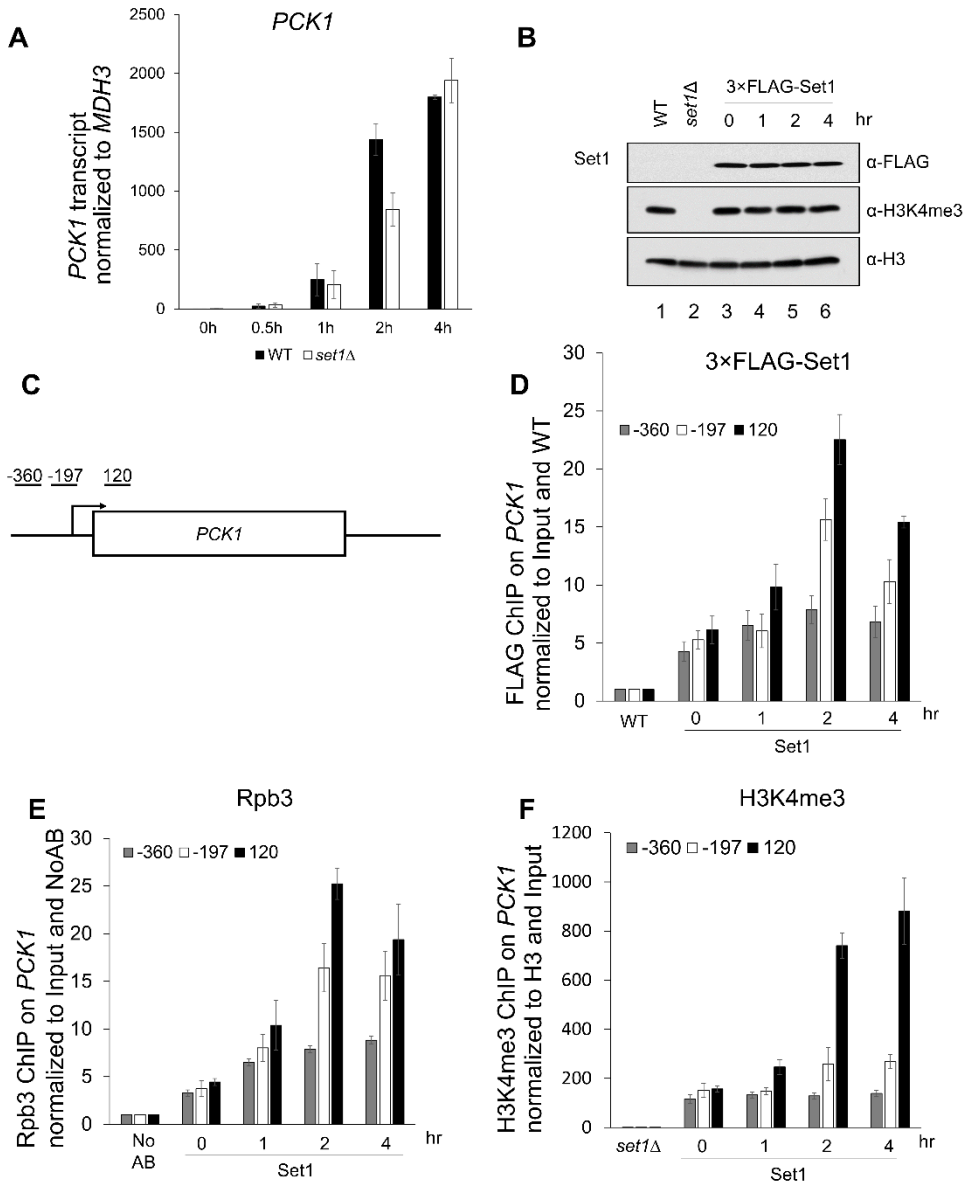


Figure 2.6 Set1 is recruited to the *PCK1* gene to deposit H3K4 trimethylation after the switch of yeast growth media from glucose to acetate.

(A) The transcript levels of *PCK1* is increasing in both WT and *set1*Δ strain after yeast media switched from glucose to acetate. The relative transcript levels of *PCK1* are determined by qRT-PCR analysis. Indicated yeast strain was collected before the switch and 0.5, 1, 2 and 4 hours after the switch from SC glucose to SC acetate. Data was normalized to transcript levels of *MDH3* and relative to untagged WT collected before the media switch. (B) The global Set1 protein levels and global H3K4 trimethylation is not changed after yeast media switched from glucose to acetate for 4 hours. Indicated yeast strain was collected before the switch and 1, 2 and 4 hours after the switch from SC containing glucose to SC containing acetate, and the whole cell

extracts were analyzed by Western blot using M2 α-FLAG and α-H3K4me3 antibodies. The whole cell lysates were blotted with α-H3 for input levels. (C) Schematic of regions analyzed for ChIP using specific primers across the *PCK1* gene. (D) ChIP analysis of 3×FLAG tagged Set1 was performed using M2 α-FLAG antibody, and the signal is normalized to input and relative to WT. Sample collection is same with Figure 2.6B. (E) ChIP analysis of Rpb3 recruitment on *PCK1* gene was performed using α-Rpb3 antibody. The signal is normalized to input and relative to WT with beads with no antibody. Sample collection is same with Figure 2.6B. (F) ChIP analysis of *3×FLAG-SET1* strain was performed using α-H3K4me3 antibody, and the signal is normalized to H3 and input levels, then relative to the signal of *set1*Δ strain. Sample collection is same with Figure 2.6B. Three technical repeats were performed for each of the three biological repeats in all ChIP and Gene expression analysis in Figure 2.6 and the error bars represent the standard error of the mean.

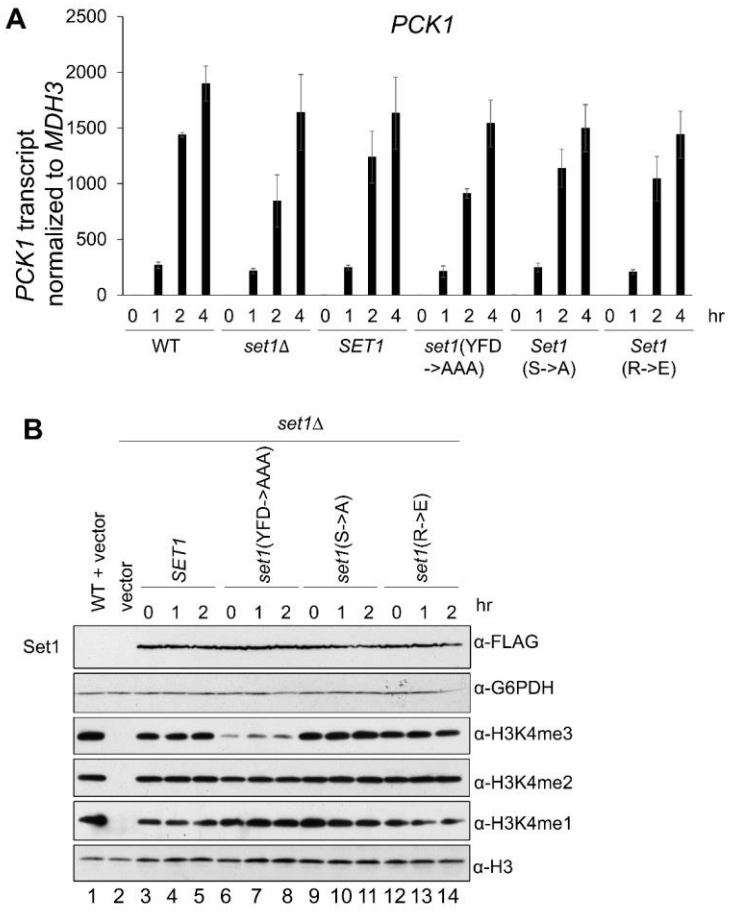


Figure 2.7 The set1 mutants do not affect transcript levels of *PCK1* and the Set1 protein levels and global H3K4 methylation levels are not changed after the switch of yeast growth media from glucose to acetate.

(A) The transcript levels of *PCK1* is increasing in set1 mutants when yeast growth media switched from glucose to acetate. The relative transcript levels of *PCK1* are determined by qRT-PCR analysis. Sample collection is same with Figure 2.6B and the media used is SC-Ura. Data was normalized to *MDH3* and relative to untagged WT collected at time 0 hour. Three technical repeats were performed for each of the three biological repeats and the error bars represent the standard error of the mean. (B) The global Set1 protein levels and global H3K4 mono-, di-, and trimethylation levels of Set1 and set1 mutants were determined after yeast growth media switching form glucose to acetate for 2 hours. Sample collection is same with Figure 2.7A. The whole cell extracts were analyzed by Western blot using M2 α-FLAG, α-H3K4me1, α-H3K4me2, and α-H3K4me3 antibodies. The whole cell lysates were blotted with α-H3 and α-G6PDH for input levels.

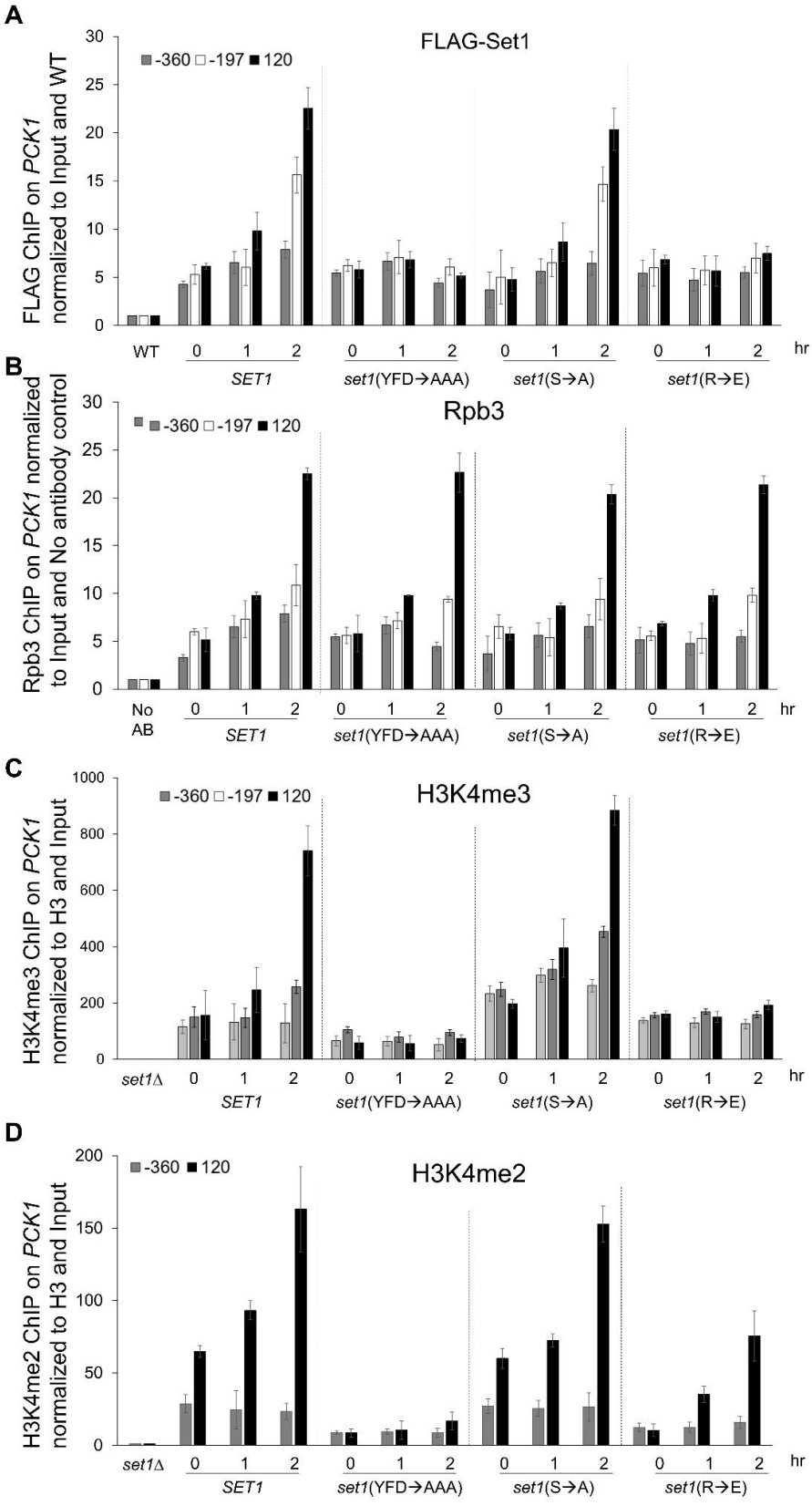

A
FLAG-Set1
-360 -197 120
FLAG ChIP on PCK1 normalized to Input and WT
WT
SET1
set1(YFD→AAA)
set1(S→A)
set1(R→E)
hr
B
Rpb3
Rpb3 ChIP on PCK1 normalized to Input and No antibody control
No AB
C
H3K4me3
H3K4me3 ChIP on PCK1 normalized to H3 and Input
set1Δ
D
H3K4me2
H3K4me2 ChIP on PCK1 normalized to H3 and Input

Figure 2.8 The SRIM is required for Set1 recruitment by RNAPII and H3K4 trimethylation establishment at 5' of induced *PCK1* gene

(A) ChIP analysis of 3×FLAG tagged Set1 recruitment on *PCK1* gene in indicated set1 mutants was performed using M2 α-FLAG antibody, and the signal is normalized to input and relative to WT. Indicated yeast strain was collected at before the switch and 1 and 2 hours after the switch of media from SC-Ura to SC-Ura acetate. (B) ChIP analysis of Rpb3 recruitment on *PCK1* gene in indicated set1 mutants was performed using α-Rpb3 antibody. The signal is normalized to input and relative to WT with beads but no antibody. Sample collection is same with Figure 2.7A. (C) ChIP analysis of H3K4 trimethylation establishment on *PCK1* gene in indicated set1 mutants was performed using α-H3K4me3 antibody, and the signal is normalized to H3 and input levels, then relative to the signal of *set1*Δ strain. Sample collection is same with Figure 2.7A. Three technical repeats were performed for each of the three biological repeats in all ChIP analysis in Figure 2.8 and the error bars represent the standard error of the mean. (D) ChIP analysis of H3K4 dimethylation establishment on *PCK1* gene in indicated set1 mutants was performed using α-H3K4me2 antibody, and the signal is normalized to H3 and input levels, then relative to the signal of *set1*Δ strain. Sample collection is same with Figure 2.7A. Three technical repeats were performed for each of the three biological repeats in all ChIP analysis in Figure 2.8 and the error bars represent the standard error of the mean.

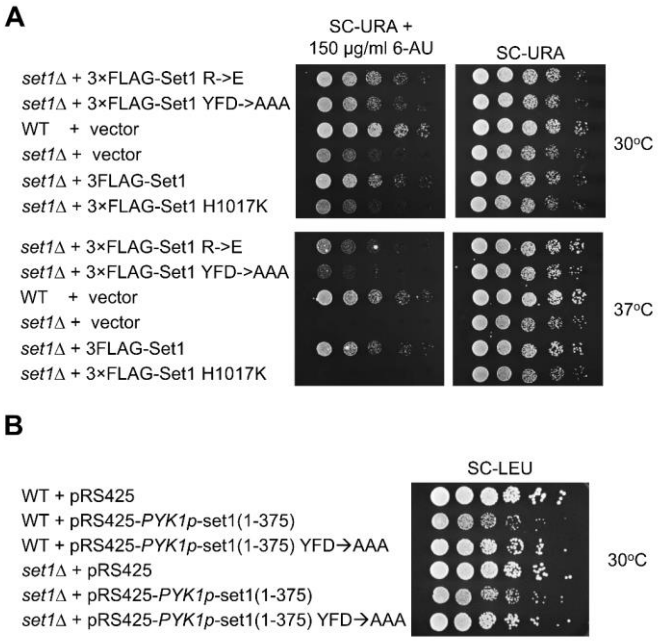


Figure 2.9 set1 SRIM mutants exhibit sensitivity on 6-AU plate and overexpression of set1(1-375) shows a slight grow defect

(A) Set1 and set1 mutants were grown to log phase in SC-Ura media and spotted in 5-fold dilutions from an OD600 0.1 on plates with or without 150 mg/L 6AU. Plates were grown at 30ºC and 37ºC. (B) Indicated log phase strains were spotted in 5-fold dilutions from an OD600 0.1 on SC-Leu plate and grown at 30ºC.

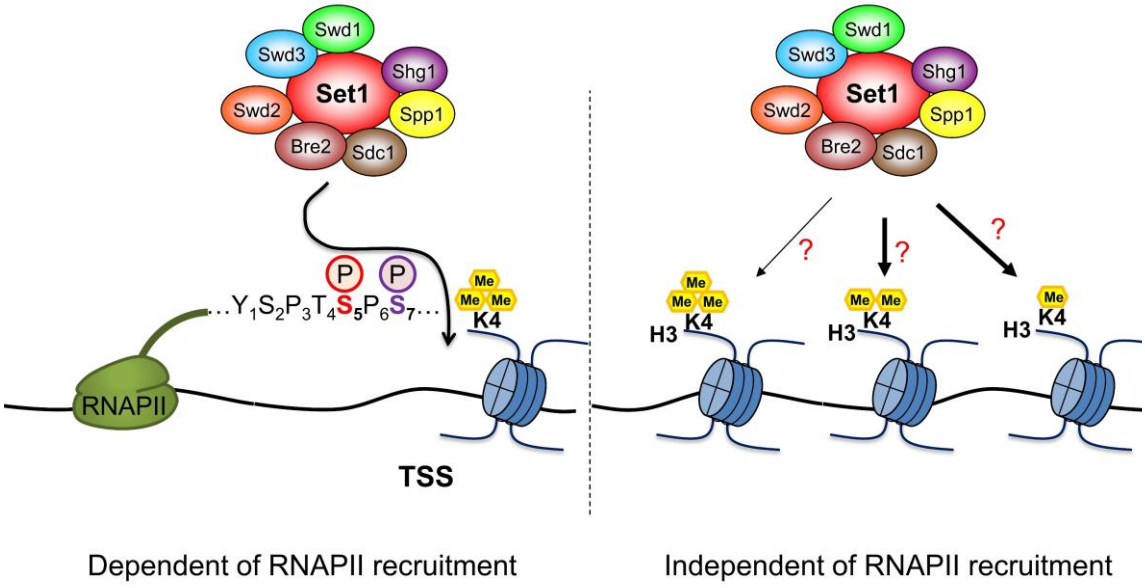


Figure 2.10 Working model

Set1 can be recruited to chromatin to deposit H3K4 trimethylation at transcription start site through a RNAII recruitment manner. Set1 can also be recruitment to chromatin independent of RNAPII through unknown mechanisms to deposit H3K4 methylation.

CHAPTER 3. N-ICE PLASMIDS FOR GENERATING N-TERMINAL 3×FLAG TAGGED GENES THAT ALLOW INDUCIBLE, CONSTITUTIVE, OR ENDOGENOUS EXPRESSION IN *SACCHAROMYCES CEREVISIAE*

3.1 Declaration of collaborative work

In this study, gene expression analysis of *SET2*, *PKC1*, and *ERG11* and West blotting analysis of *PKC1* and *ERG11* were performed by Nina D. Serratore. All the other experiments and analysis were performed by Yueping Zhang.

3.2 Introduction

Saccharomyces cerevisiae (*S. cerevisiae*) is highly efficient in homologous recombination, which renders PCR-mediated gene modifications such as gene deletions and epitope tagging possible (171, 172). Epitope tagging of genes is one of the most common strategies to facilitate protein detection, purification, and localization in cells. PCR-mediated gene modification can integrate epitope tags at the N-terminus or C-terminus of target genes in yeast. A plasmid containing a selection marker and the epitope tag of interest can be used as the template for PCR amplification of a DNA cassette for genomic integration (173). The PCR primer contains 45-60 additional nucleotides that are homologous to upstream and downstream sequences of the integration site in the genome. By transforming yeast with the amplified PCR cassette

with flanking sequences homologous to the gene of interest, the selection marker cassette with the epitope tag will be integrated.

C-terminal epitope tagging is relatively easy to perform because only one transformation step is required (174-176). The expression levels of a C-terminal tagged gene is comparable to that of the endogenous untagged gene because both are expressed from the same promoter.

However, C-terminal epitope tagging may not be feasible for all proteins. Some proteins, such as Smt3, are processed at the C-terminus, resulting in a loss of the epitope tag after cleavage (177). Whereas other proteins, such as the Set1 histone H3K4 methyltransferase, a C-terminal epitope tag renders the protein nonfunctional (109, 169, 178).

Under these circumstances, the N-terminal epitope tagging system is needed. Because integrating a selection marker at the N-terminus of the target gene will disrupt expression, different systems have been developed. One way to avoid this problem is to incorporate a heterologous promoter (e.g. *ADH1*, *CYC1*, *GPD*, *TEF*, and *GAL1*) in integration cassettes upstream of the epitope tag (174). Although, the tagged protein is not expressed at the endogenous level and the selection marker is not removed. Another method is to excise the selection marker after the integration, leaving the tagged protein expressed from the endogenous promoter (179, 180). However, N-terminal tagging essential genes in haploid yeast is a problem due to the disruption of the endogenous promoter by the drug resistant marker.

Many different epitopes are available in *S. cerevisiae*, such as 6×HA, 9×MYC, GFP, biotin, and TAP tag (174, 175, 179, 181-183). Multiple repeats of the HA and MYC

epitope tags are used for enhancing the detection of lowly expressed proteins. The FLAG epitope, Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys, is used because of the availability of high affinity and specific monoclonal antibodies that recognize the FLAG sequence. Triple tandem repeats of the FLAG (3×FLAG) epitope can increase sensitivity for detection and does not affect protein function in most cases likely due to the small overall size (184). The 3×FLAG epitope is widely used for protein detection in assays such as immunoprecipitation, immunofluorescence, and protein purification in multiple model systems. Currently, the N-terminal 3×FLAG epitope tagging system does not readily permit tagging of essential genes under endogenous expression when using haploid yeast (185, 186).

In this study, we have constructed the N-ICE plasmid system for PCR-based N-terminal 3×FLAG epitope tagging of non-essential and essential genes under the control an Inducible promoter (*GAL1*), Constitutive promoters (*CYC1* or *PYK1*), or the Endogenous promoter. Our N-ICE plasmid system also provides the option of three different selection markers, including *KanMX*, *HphMX*, and *NatMX* cassettes. In addition, the integrated selection marker and the heterologous promoter can be excised with a subsequent step using Cre recombinase, leaving the target gene expressed from the endogenous promoter. This system can also be applied to tagging essential genes expressed by their endogenous promoters. We have validated the N-ICE system by N-terminal tagging two non-essential genes (*SET1* and *SET2*) and two essential genes (*ERG11* and *PKC1*). Overall, the N-ICE plasmid system will provide valuable new tools to enhance the yeast epitope tagging toolbox.

3.3 Results

In this study, the N-ICE plasmid system was generated for N-terminal 3×FLAG epitope tagging of non-essential and essential genes under the control of an Inducible, Constitutive, or Endogenous promoter. The first generation N-ICE plasmids include pK3F, pK3F-*PYK1*, pK3F-*GAL1*, pH3F, pH3F-*PYK1*, and pH3F-*GAL1* (Figure 3.1 A-F Figure 3.2 A-F). To avoid undesired PCR amplification, sequences were designed to use different codons for the same amino acids to avoid repeats of DNA sequences in the 3×FLAG epitope tag. The first construct, pK3F, was generated with a *loxP-KanMX-loxP* cassette followed by a 3×FLAG epitope sequence (Figure 3.1A). In the plasmids pK3F-*PYK1* and pK3F-*GAL1*, the *PYK1* promoter (-811bp to -1bp) and the *GAL1* promoter (-457bp to -5bp) were inserted into the pK3F plasmid downstream of the *KanMX* cassette and upstream of the second loxP site, respectively (Figure 3.1B-C). The pH3F, pH3F-*PYK1* and pH3F-*GAL1* vectors were generated by replacing the *KanMX* cassettes in pK3F, pK3F-*PYK1* and pK3F-*GAL1* with *HphMX* cassettes. The P1 and P2 sites were designed for PCR amplification of the integration cassette containing the 3×FLAG epitope tag (Figure 3.3A and B). For broader application and compatibility with previous N-terminal tagging systems, 8 additional second generation N-ICE plasmids were created. Second generation N-ICE plasmids include pK3FS, pK3FS-*PYK1*, pK3FS-*GAL1*, pK3FS-*CYC1*, pH3FS, pH3FS-*PYK1*, pH3FS-*GAL1*, and pN3FS-*CYC1* (Figure 3.1G-N and Figure 3.2 G-N). All the second generation plasmids contain additional S1 and S4 sites flanking the previous P1 and P2 sites (Figure 3.3A and B). The S1 and S4 sites are commonly used as primer sets by other N-terminal tagging systems (174, 187). In order to prevent low tagging efficiency due to the replacement of an existing selection

marker, the terminators of the *HphMX* cassette and the *NatMX* cassette have been altered in the second generation N-ICE plasmids (see details in the Materials and methods section). In addition, the *CYC1* promoter (-287bp to -1bp) was used as the heterologous promoter in pK3FS-*CYC1* and pN3FS-*CYC1*, since some genes will cause lethality when overexpressed from a strong promoter (188, 189).

The N-ICE strategy for PCR epitope tagging is presented in Figure 3.3A and described in the methods. Briefly, the tagging cassettes are PCR amplified from plasmid templates using primers containing sequences homologous to the targeted gene. Depending on the construct used, the 3×FLAG-tagged genes will be expressed from a strong constitutive *PYK1* promoter, a weak constitutive *CYC1* promoter, or an inducible *GAL1* promoter (Figure 3.3A). By transforming pSH47 and expressing Cre recombinase under galactose growth conditions, the integrated selection marker and promoter between the two loxP sites are excised, leaving the 3×FLAG epitope tagged target gene with the endogenous promoter and a loxP site (Figure 3.3A).

The pK3F construct was validated by N-terminal 3×FLAG tagging Set1. The pK3F cassette was PCR amplified with Set1 primers (*3×FLAG-SET1*-F and *3×FLAG-SET1*-R) and genomic integration was validated using the *SET1*-271F and *KAN*-001R primers (Table 3.1 and data not shown). To determine if the 3×FLAG epitope tag and the remaining loxP site affected the transcript levels of *SET1*, quantitative real-time PCR (qRT-PCR) analysis was performed (Figure 3.4A). Transcript levels of *SET1* were normalized to Actin (*ACT1*) and a WT untagged strain. Importantly, no significant transcript differences are detected between the WT strain and the *3×FLAG-SET1* strain. Western blot analysis using the monoclonal M2 FLAG antibody showed that the tagged

Set1 protein is expressed and detected in the *3×FLAG-SET1* strain, but not in WT or *set1Δ* strains (Figure 3.4B). Because Set1 is a histone H3K4 methyltransferase whose activity is inhibited when tagged at the C terminus, but not N-terminus (109, 169, 178), H3K4 methylation levels were analyzed (Figure 3.4B). H3K4 trimethylation levels in the *3×FLAG-SET1* strain is similar to that of a WT strain, indicating that the 3×FLAG tag did not disrupt or alter the methyltransferase activity of Set1. Histone H3 was used as a loading control.

To validate the pK3F-*PYK1* and pK3F-*GAL1* N-ICE constructs, the Set2 histone H3K36 methyltransferase was N-terminal 3×FLAG-tagged. The pK3F, pK3F-*PYK1*, and pK3F-*GAL1* cassettes were PCR amplified with Set2 primers (*3×FLAG-SET2*-F and *3×FLAG-SET2*-R) and the integrations were validated using the *SET2*-001F and *KAN*-001R primers (Table 3.1 and data not shown). To determine transcript levels of *SET2* under different promoters, quantitative real-time PCR (qRT-PCR) analysis was performed (Figure 3.4C). Transcript levels of *SET2* were normalized to Actin (*ACT1*) and set relative to a WT untagged strain. Transcript levels of *3×FLAG-SET2* driven from the *PYK1* promoter and the induced *GAL1* promoter are significantly higher than the WT strain and the endogenously tagged *3×FLAG-SET2* strain (Figure 3.4C). Additionally, the *SET2* transcript level is repressed under the *GAL1* promoter when grown under glucose conditions compared to the WT untagged strain (Figure 3.4C). Immunoblots using the monoclonal M2 FLAG antibody determined that Set2 is highly expressed from the constitutive *PYK1* promoter and when the *GAL1* promoter is induced (Figure 3.4D, lanes 3 and 5). The immunoreactive bands are specific for the 3×FLAG-tagged Set2 since Set2 is not detected in the WT strain, the strain containing the integrated pK3F cassette prior

to Cre excision, or the strain containing the integrated pK3F-*GAL1* cassette under uninduced growth conditions (Figure 3.4D, lanes 1, 2, and 4). The asterisks mark likely protein degradation products of Set2 when Set2 is overexpressed. Importantly, the removal of the *KanMX* selection marker by Cre recombinase showed that all the 3×FLAG-tagged Set2 strains are expressed at similar levels regardless of whether they were initially tagging using pK3F, pK3F-*PYK1*, and pK3F-*GAL1* vectors (Figure 3.4D, lanes 6-8, and Figure 3.7A). G6PDH was used as a loading control. Set2 is the only histone H3K36 methyltransferase in *S. cerevisiae*, and the deletion of *SET2* abolishes all H3K36 methylation (190, 191). To determine whether the 3×FLAG-Set2 is still functional in yeast, the *in vivo* histone H3K36 trimethylation levels were determined by Western blot (Figure 3.4E). As expected, the H3K36 trimethylation is abolished in the integrated pK3F transformed cells prior to excision of loxP sites since *SET2* expression is abolished (Figure 3.4E, lane 2 and Figure 3.4D, lane 2). The H3K36 trimethylation level is also abolished in the *GAL1p-3×FLAG-SET2* containing strain when grown under glucose-repressed conditions, but is recovered in galactose-inducible conditions (Figure 3.4E, lane 4 versus lane 5). There is no loss of H3K36 trimethylation using the pK3F-*PYK1* (Figure 3.4E, lane 3). In addition, the removal of the *KanMX* selection marker by Cre recombinase showed that all the 3×FLAG-tagged *SET2* strains had H3K36 trimethylation levels similar to a WT strain (Figure 3.4E, lanes 6-8). H3 was used as a loading control.

To determine if the N-ICE plasmid system can be used to tag essential genes, *PKC1* and *ERG11* were N-terminal 3×FLAG-tagged. The *PKC1* gene encodes a homolog of the mammalian protein kinase C (PKC) and is essential for yeast cell growth (192).

The *ERG11* gene encodes a cytochrome P450 lanosterol demethylase and is essential for the ergosterol biosynthesis pathway and cell viability (193). The *PKC1* and *ERG11* integration cassettes were amplified from the pH3F-*PYK1* construct using the *3×FLAG-PKC1*-F and *3×FLAG-PKC1*-R or *3×FLAG-ERG11*-F and *3×FLAG-ERG11*-R primers (P1 and P2 primer sets), and the integrations were validated using the *PKC1*-001F and *HYG*-001R primers, or by using the *ERG11*-001F and *HYG*-001R primers, respectively (Table 3.1 and data not shown). To determine the transcript levels of *PKC1* and *ERG11* under the *PYK1* promoter or the endogenous promoter, quantitative real-time PCR (qRT-PCR) analysis was performed (Figure 3.5A and B). Transcript levels of *PKC1* and *ERG11* were normalized to Actin (*ACT1*) and relative to a WT untagged strain. The transcript levels of *3×FLAG-PKC1* and *3×FLAG-ERG11* from the *PYK1* promoter are significantly higher than the WT strain (Figure 3.5A and B). *PKC1* and *ERG11* transcript levels from the *3×FLAG-PKC1* and *3×FLAG-ERG11* strains, respectively, are not significantly different compared to the WT untagged strain (Figure 3.5A and B). This indicates the insertion of 3×FLAG epitope tag does not affect the transcript levels of the target gene. Immunoblots using the monoclonal M2 FLAG antibody determined that Pkc1 is overexpressed from the constitutive *PYK1* promoter (Figure 3.5C, lanes 1 and 2) when compared to endogenous expression (Figure 3.5C, lanes 3 and 4, and Figure 3.7B). The Erg11 expression under the *PYK1* promoter is increased about two fold compared to the expression under the endogenous promoter (Figure 3.5D, lanes 1 and 2 vs. 3 and 4, and Figure 3.7C). A serial dilution spot assay showed that overexpression of 3×FLAG-Erg11 or 3×FLAG-Pkc1 under the control of the *PYK1* promoter did not affect yeast viability and grew similar to the WT strain (Figure 3.5E).

In order to test whether our second generation N-ICE plasmids can be applied to tag essential genes, the *PKC1* gene was N-terminal 3×FLAG-tagged using the S1 and S4 primer set. Integration cassettes were PCR amplified from the following plasmids pK3FS-*PYK1*, pK3FS-*CYC1*, pH3FS-*PYK1*, and pN3FS-*CYC1* using the *3×FLAG-PKC1*-S1-F and *3×FLAG-PKC1*-S4-R primers. The integrated cassettes were validated by PCR (data not shown). To determine the transcript levels of *PKC1* under different promoters, the qRT-PCR analysis was performed. The transcript levels of *PKC1* expressed from the *PYK1* promoter (pK3FS-*PYK1* and pH3FS-*PYK1*) are significantly higher than the untagged WT strain (Figure 3.6A). The transcript levels of *PKC1* expressed from *CYC1* promoters (pK3FS-*CYC1* and pN3FS-*CYC1*) are about three fold higher than the untagged WT strain (Figure 3.6B). The transcript levels of *3×FLAG-PKC1* expressed from the endogenous promoter are not significantly different than the untagged WT strain (Figure 3.6A and B). Immunoblots using the monoclonal M2 FLAG antibody determined that the Pkc1 protein levels from the *PYK1* promoter (Igure 3.6C and D, lanes 1 and 2 and Figure 3.7D and E) are significantly higher than Pkc1 expressed endogenously (Figure 3.6C and D, lanes 5 and 6 and Figure 3.7D and E). As expected, the protein level of Pkc1 expressed from the weak constitutive *CYC1* promoter is higher than Pkc1 expressed from the endogenous promoter but lower than Pkc1 expressed from the strong constitutive *PYK1* promoter (Figure 3.6C and D, and Figure 3.7D and E).

3.4 Discussion

We have generated the N-ICE plasmid system for generating N-terminal 3×FLAG epitope tagged essential and non-essential genes. These new N-ICE plasmids allow the tagged gene of interest to be expressed from the endogenous promoter or overexpressed

from a constitutive *PYK1*, *CYC1* or inducible *GAL1* promoter. Because N-ICE plasmids can be readily modified, the N-ICE system can be further diversified with different selection markers, promoters, or epitope tags. Overall, we have provided a new and valuable N-ICE plasmid system for the yeast community and expanded the toolbox for epitope tagging.

3.5 Materials and methods

Strains and growth media

All yeast strains were derived from *S. cerevisiae* BY4741 (*MATa his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*). Yeast cells were grown at 30 °C in YPD medium (1% yeast extract, 2% peptone, 2% glucose) or synthetic complete (SC) medium with amino acid dropouts (194). Geneticin (G418, GoldBio) was used at the final concentration of 200 μg/ml for antibiotic selection of the *KanMX* cassette. Hygromycin B (Gibico, Thermo Fisher Scientific) was used at the final concentration of 300 μg/ml for antibiotic selection of the *HphMX* cassette. Nourseothricin Sulfate (GoldBio) was used at the final concentration of 100 μg/ml for antibiotic selection of the *NatMX* cassette.

Plasmid construction

Oligonucleotides encoding the 3×FLAG epitope sequence, 3×FLAG-T and 3×FLAG-B (Table 3.1), were annealed, generating overhanging ends compatible with SalI and BamHI. The annealed oligonucleotides were ligated into the SalI and BamHI sites of pUG6-myc-N-Avitag (183) to generate pUG6-3×FLAG. The *KanMX* cassette with the *Ashbya gossypii TEF* (*AgTEF*) promoter and *AgTEF* terminator was amplified from the pUG6-myc-N-Avitag plasmid using the *KAN*-1F and *KAN*-2R primers. The *HphMX* cassette with the *AgTEF* promoter and *AgTEF* terminator were amplified from

the pAG32 plasmid (168) also using the *KAN*-1F and *KAN*-2R primers. The PCR amplification product was purified and digested with SalI and BglII, and ligated into XhoI and BglII sites of pUG6-3×FLAG to generate the pK3F and pH3F plasmids. The *PYK1* and *GAL1* promoters were PCR amplified from yeast BY4741 genomic DNA, using the *PYK1*-F and *PYK1*-R or the *GAL1*-F and *GAL1*-R primers, respectively. The PCR products were digested with SpeI and XbaI and ligated into the XbaI site of the pK3F or pH3F plasmids to generate the pK3F-*PYK1* and pK3F-*GAL1*, or pH3F-*PYK1* and pH3F-*GAL1* plasmids, respectively.

To incorporate the S1 and S4 primer sets in our tagging system, integration cassettes were amplified using primers containing S1 and S4 sites and ligated to the plasmid backbone. To generate pK3FS, pK3FS-*PYK1*, and pK3FS-*GAL1*, integration cassettes from pK3F, pK3F-*PYK1*, and pK3F-*GAL1* were amplified using S1-001F and S4-003R primer sets. The PCR products were digested with NotI and XhoI and ligated in the NotI and XhoI sites of the pUG6 PCR product. The pUG6 PCR product was generated from the pUG6-myc-N-Avitag using pUG-003F and pUG-002R primer sets. To generate the pH3FS, pH3FS-*PYK1*, and pH3FS-*GAL1* vectors and switch the terminator to the *Ashbya gossypii ADH1* (*AgADH1*) terminator, the *HphMX* cassette was amplified by S1-001F and *HphMX*-001R from pH3F, digested with NotI and BglII, and ligated in the NotI and BglII sites of pK3FS, pK3FS-*PYK1*, and pK3FS-*GAL1*. To generate the pK3FS-*CYC1* vector, the *CYC1* promoter was PCR amplified using *CYC1*-001F and *CYC1*-002R from *Saccharomyces cerevisiae* genomic DNA, digested with BglII and XbaI, and ligated in the BglII and XbaI sites of pK3FS-*PYK1*. The *NatMX* cassette with the *Saccharomyces cerevisiae ADH1* (*ScADH1*) terminator was amplified

from the pAG25 plasmid using the *KAN*-1F and *NatMX*-001R primers (168). The PCR amplification product was purified and digested with SalI and BglII, and ligated into XhoI and BglII sites of pUG6-3×FLAG to generate the pN3F plasmid. To generate the pN3FS-*CYC1* vector, the *NatMX* cassette was amplified by S1-001F and *NatMX*-001R from pN3F and digested with NotI and BglII, and ligated in the NotI and BglII sites of pK3FS-*CYC1*. All constructs were verified by PCR and DNA sequencing.

PCR amplification of the N-terminal 3×FLAG-tagging cassette for genomic integration

The PCR conditions used were denaturation at 95 °C for 5 min; 95 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min for a total of 30 cycles, followed by a final elongation step at 72 °C for 10 min. The final PCR product was purified from an agarose gel and ethanol precipitation.

Yeast transformation and excision of the integration cassette

Yeast strain BY4741 was transformed with the purified PCR product and selected on YPD plates with designated antibiotics. The colonies were verified by genomic PCR amplification (195). Once verified, the pSH47 plasmid containing the Cre recombinase under the *GAL1* promoter was transformed into the N-terminal 3×FLAG tagged strain and selected on a SC-Ura plate (180). The colonies were streaked on SC-Ura plates. A single colony was inoculated into 5 mL of SC-Ura media and grown at 30 °C overnight. To induce Cre expression, 5 μL of the overnight culture was diluted into 5 mL of SC-Ura media containing 2% galactose (the sole carbon source) and incubated for 5 hours at 30 °C with agitation at 250rpm. The Cre recombinase expressed cells (5 μl) were mixed with 100 μL of water and plated onto YPD plates. To verify the selection marker

excision, colonies were streaked onto both a YPD plate and a YPD plate with designated antibiotics. The colonies that grew on YPD plates, but not the YPD plates containing antibiotics, were selected for PCR verification. PCR validated strains were streaked onto a 5-FOA plate to counter select for the pSH47 Cre plasmid.

Gene expression analysis

RNA transcript levels were determined using quantitative real-time PCR (qRT-PCR) as previously described (133). Three technical repeats were performed for each of the three biological repeats. All ΔΔCT values were analyzed by normalizing values to *ACT1* and setting the transcript levels of targeted genes relative to the transcript level in an untagged WT strain. The primer sequences for *SET1*, *SET2*, *PKC1*, *ERG11*, and *ACT1* are found in Table 3.1.

Western blot analysis and antibodies used

Yeast extraction and Western blot analysis to detect the 3×FLAG tagged proteins were performed as described previously (113, 119, 169). The α-FLAG mouse antibody (Sigma, M2) was used at 1:5,000. The glucose-6-phosphate dehydrogenase (G6PDH) rabbit antibody (Sigma, A9521) was used at 1:50,000. The α-H3K36me3 rabbit antibody (Abcam, Ab9050) was used at 1:5,000. The α-H3 rabbit antibody (Active Motif, 39163) was used at 1:5,000. The α-H3K4me3 rabbit antibody (Active Motif, 39159) was used at 1:100,000. The Western blots were quantified using ImageJ.

Table 3.1 Primer Sequences in Chapter 3

Name	Sequence
3×FLAG-T	5’-TCGACACCATGGACTATAAAGACGACGACGACAAAGATT ATAAGGACGACGACGATAAGGATTACAAGGACGACGACGA TAAAGGATCCATGGGCGGCCGCGA-3’
3×FLAG-B	5’-GATCTCGCGGCCGCCCATGGATCCTTTATCGTCGTCGTCCT TGTAATCCTTATCGTCGTCGTCCTTATAATCTTTGTCGTCGTC GTCTTTATAGTCCATGGTG-3’
PYK1-F	5’-GGGACTAGTCGCCCTGGTCAAACTTCAGAAC-3’
PYK1-R	5’-CCCGGATCCGGTCTAGATGTGATGATGTTTTATTTGTTTT GATTGG-3’
GAL1-F	5’-GGGACTAGTAGTACGGATTAGAAGCCGC-3’
*GAL*1-R	5’-CCCGGATCCGGTCTAGAGTTTTTTCTCCTTGACGTTAAAG TATA-3’
KAN-1F	5’-CCCCGTCGACGTCCCCGCCGGGTCACCCGGCCAG-3’
KAN-2R	5’-GGCCAGATCTCGACACTGGATGGCGGCGTTAGTATC-3’
S1-001F	5’-CAGATCCGCGGCCGCGTACGCTGCAGGTCGACGGATCTGA TATCACCTAATAAC-3’
S4-003R	5’- CGATCTCGAGCATCGATGAATTCTCTGTCGCCATGGATCCT TTATCGTCG -3’
pUG-003F	5’-CTAGCTCGAGGTGCACTCTCAGTACAATCTGCTC-3’
pUG-002R	5’-CCTATGCGGCCGCGGATCTG-3’
HphMX-001R	5’- GCCGAGATCTAAAAGTCAGGCGATCACACATTAAAAGCTA TACATTACAAAGCGTTAATAGCGGGCATTATTCCTTTGCCCTC GGAC -3’
NatMX-001R	5’- GCGCAGATCTATACACTTATTTTTTTTATAACTTATTTAATA ATAAAAATCATAAATCATAAGAAATTCGCTTAGGGGCAGGGCATGCT C -3’
CYC1-001F	5’- GCGCAGATCTGCGTTGGTTGGTGGATCAAGC -3’
CYC1-002R	5’- GCGCTCTAGATATTAATTTAGTGTGTGTATTTGTGTTTGTGTG-3’
3×FLAG-SET1-F	5’-TCCTTATTTGTTGAATCTTTATAAGAGGTCTCTGCGTTTAG AGA**GGATCTGATATCACCTAA**-3’
3×FLAG-SET1-R	5’-TCTGTATGAACCAGAAGACGCGTGTGCTCTTCTATAGTAA TTTGA**CATGGATCCTTTATCGTC**-3’
3×FLAG-SET2-F	5’-TGCTGTCAAACCTTTCTCCTTTCCTGGTTGTTGTTTTACGT GATC**GGATCTGATATCACCTAA**-3’
3×FLAG- SET2-R	5’-TATTTCTTTTTCATCTTCCGACGCACTCACACTTTGGTTCT TCGA**CATGGATCCTTTATCGTC**-3’
3×FLAG-PKC1-F	5’-GAAAAGTAAGTATAGTATCACACATATAGGGAGCAGTTTA CAGTC**GGATCTGATATCACCTAA**-3’
3×FLAG- PKC1-R	5’-GACGGCTATCTTTTTTTTAATGTTCTGCTCCAATTGTGAAA AACT**CATGGATCCTTTATCGTC**-3’

Table 3.1 continued

3×FLAG-ERG11-F	5’-TTGCAGCAGGCTTGAATAGAAACAGAACAAACGAGTAATACAAGG**GGATCTGATATCACCTAA**-3’
3×FLAG-ERG11-R	5’-GTTTACGTATTCCAATGCCTCTCCAACGATTGACTTGGTAGCAGA**CATGGATCCTTTATCGTC**-3’
3×FLAG-PKC1-S1-F	5’- AAGTAAGTATAGTATCACACATATAGGGAGCAGTTTACAGTCCGTACGCTGCAGGTCGAC -3’
3×FLAG-PKC1-S4-R	5’- CTATCTTTTTTTTAATGTTCTGCTCCAATTGTGAAAAACTCATCGATGAATTCTCTGTCG -3’
SET1-001F (GE)	5’-CACACGCGTCTTCTGGTTCAT-3’
SET1-002R (GE)	5’-TGATAGTGACCAGAACGCGAAT-3’
SET2-001F (GE)	5’-ACGTCGCAGCCAGCAAA-3’
SET2-002R (GE)	5’-TCTCCCAGCCTGGAGGTAATC-3’
ACT1-001F (GE)	5’-TGGATTCCGGTGATGGTGTT-3’
ACT1-002R (GE)	5’-TCAAAATGGCGTGAGGTAGAGA-3’
KAN-001R (V)	5’-ATCGCAGTGGTGAGTAACCAT-3’
HYG-001R (V)	5’-ATCAGCAATCGCGCATATG-3’
NAT-002R (V)	5’-GGCGGTGACGCGGAAGAC-3’
SET1-271F (V)	5’-GGACACTTGCGATTCTAGCT-3’
SET2-001F (V)	5’-CGTGATCAATATCACCTTTAACG-3’
PKC1-001F (V)	5’-GGCTGAGTAGAGACTGATTATTAACC-3’
ERG11-001F (V)	5’-GCCCTCCATGTGTATTCATTATATAAG-3’

*Bold and underlined sequences indicate the P1 and P2 sequences (see Figure 1)
*Underlined and not bold sequences indicate the S1 and S4 sequences
*GE: Gene Exp = primer sets used to verify mRNA transcript levels.
*V: Validation = primer sets used to verify genomic integration

Table 3.2 Yeast Strains and Genotype in in Chapter 3

Yeast Strain	Genotype	Reference
BY4741	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0*	Open Biosystems
SDBY1420	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 set1Δ::HphMX*	This study
SDBY1421	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 SET1::loxP-3×FLAG-SET1*	This study
SDBY1422	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 set2Δ:: loxP-KanMX-loxP-3×FLAG-SET2*	This study
SDBY1423	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 SET2::loxP-KanMX-*P_{PYK1}*-loxP-3×FLAG-SET2*	This study
SDBY1424	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 set2Δ::loxP-KanMX-*P_{GAL1}*-loxP-3×FLAG-SET2*	This study
SDBY1425	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 SET2::loxP-3×FLAG-SET2*	This study
SDBY1426	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 PKC1::loxP-HphMX-* P_{PYK1}*-loxP-3×FLAG-PKC1*	This study
SDBY1427	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 ERG11::loxP-HphMX-* P_{PYK1}*-loxP-3×FLAG-ERG11*	This study
SDBY1428	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 PKC1::loxP-3×FLAG-PKC1*	This study
SDBY1429	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 ERG11::loxP-3×FLAG-ERG11*	This study
SDBY1430	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 PKC1::loxP-KanMX-* P_{PYK1}*-loxP-3×FLAG-PKC1* (S1/S4)	This study
SDBY1431	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 PKC1::loxP-HphMX-* P_{PYK1}*-loxP-3×FLAG-PKC1* (S1/S4)	This study
SDBY1432	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 PKC1::loxP-KanMX-*P_{CYC1}*-loxP-3×FLAG-PKC1* (S1/S4)	This study
SDBY1433	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 PKC1::loxP-NatMX-* P_{CYC1}*-loxP -3×FLAG-PKC1* (S1/S4)	This study
SDBY1434	*MATa his3Δ leu2Δ0 LYS2 met15Δ0 ura3Δ0 PKC1::loxP-3×FLAG-PKC1* (S1/S4)	This study

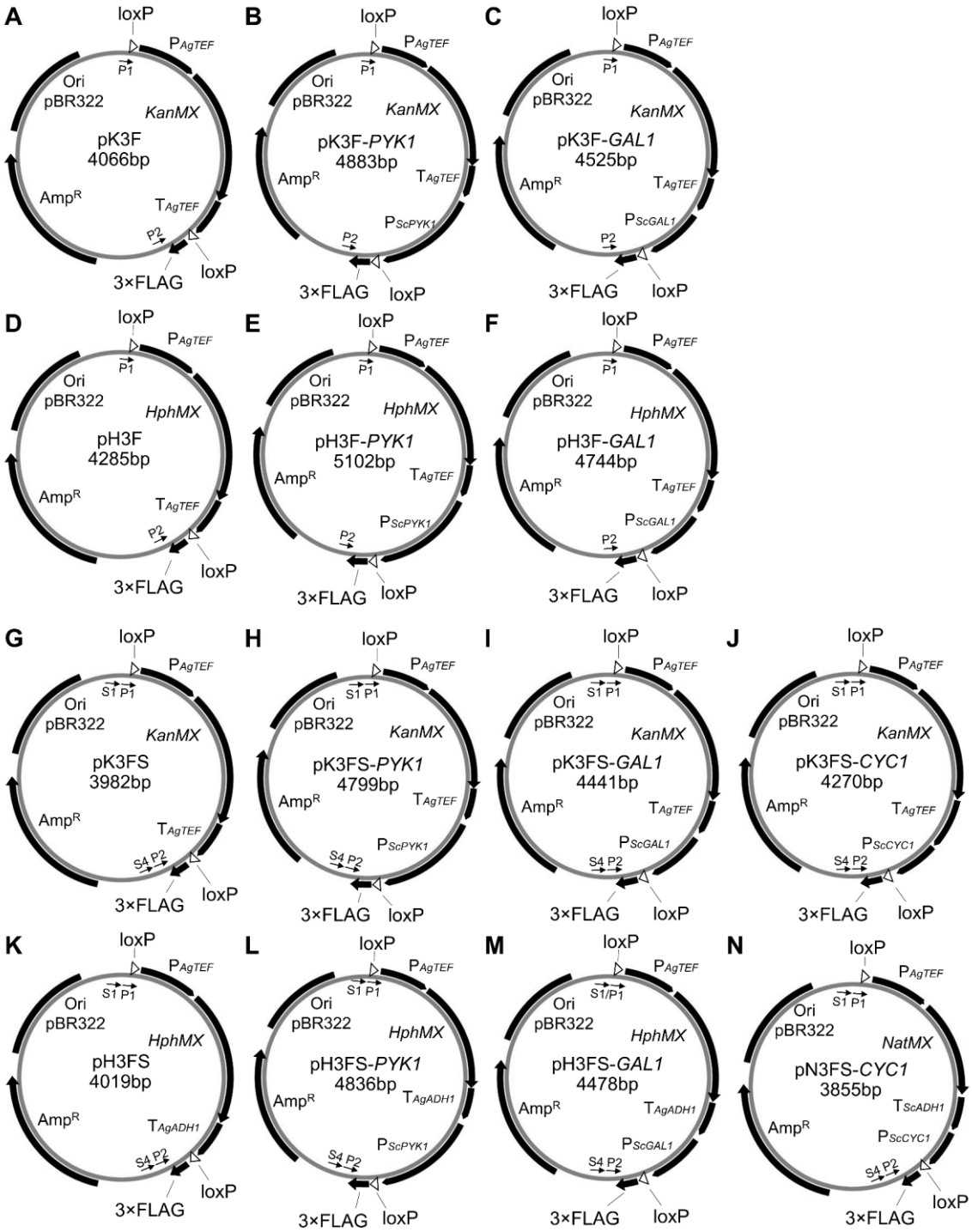


Figure 3.1 Schematic representation of N-ICE plasmids.

The 3×FLAG epitope, the *KanMX* cassette, the *HphMX* cassette, the *NatMX* cassette, heterologous promoters (*PYK1*, *GAL1*, or *CYC1*), loxP sites, P1/P2 sites, and S1/S4

sites are indicated in the figure. The *Ag* (*Ashbya gossypii*) and the *Sc* (*Saccharomyces cerevisiae*) are indicated. First generation N-ICE plasmids (A-F) have P1/P2 sequences for PCR amplification. Second genration N-ICE plasmids (G-N) have P1/P2 and S1/S4 primer sequences for PCR amplification. N-ICE plasmids pK3F, pH3F, pK3FS, and pH3FS vectors do not contain a promoter.

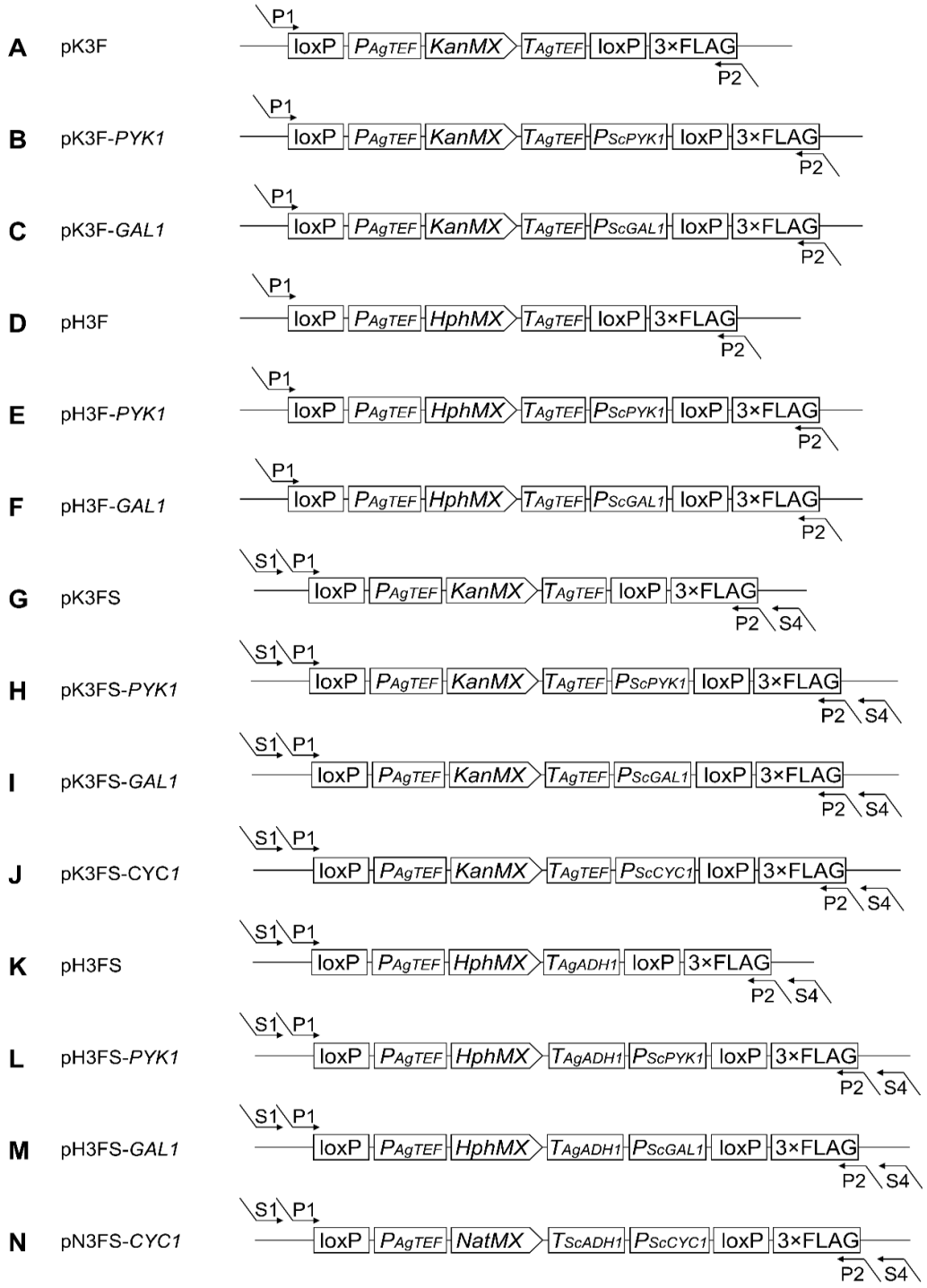


Figure 3.2 Schematic representation of integration cassettes of N-ICE plasmids.

The 3×FLAG epitope, the *KanMX* cassette, the *HphMX* cassette, the *NatMX* cassette, heterologous promoters (*PYK1*, *GAL1*, or *CYC1*), loxP sites, P1/2 and S1/S4 are indicated.

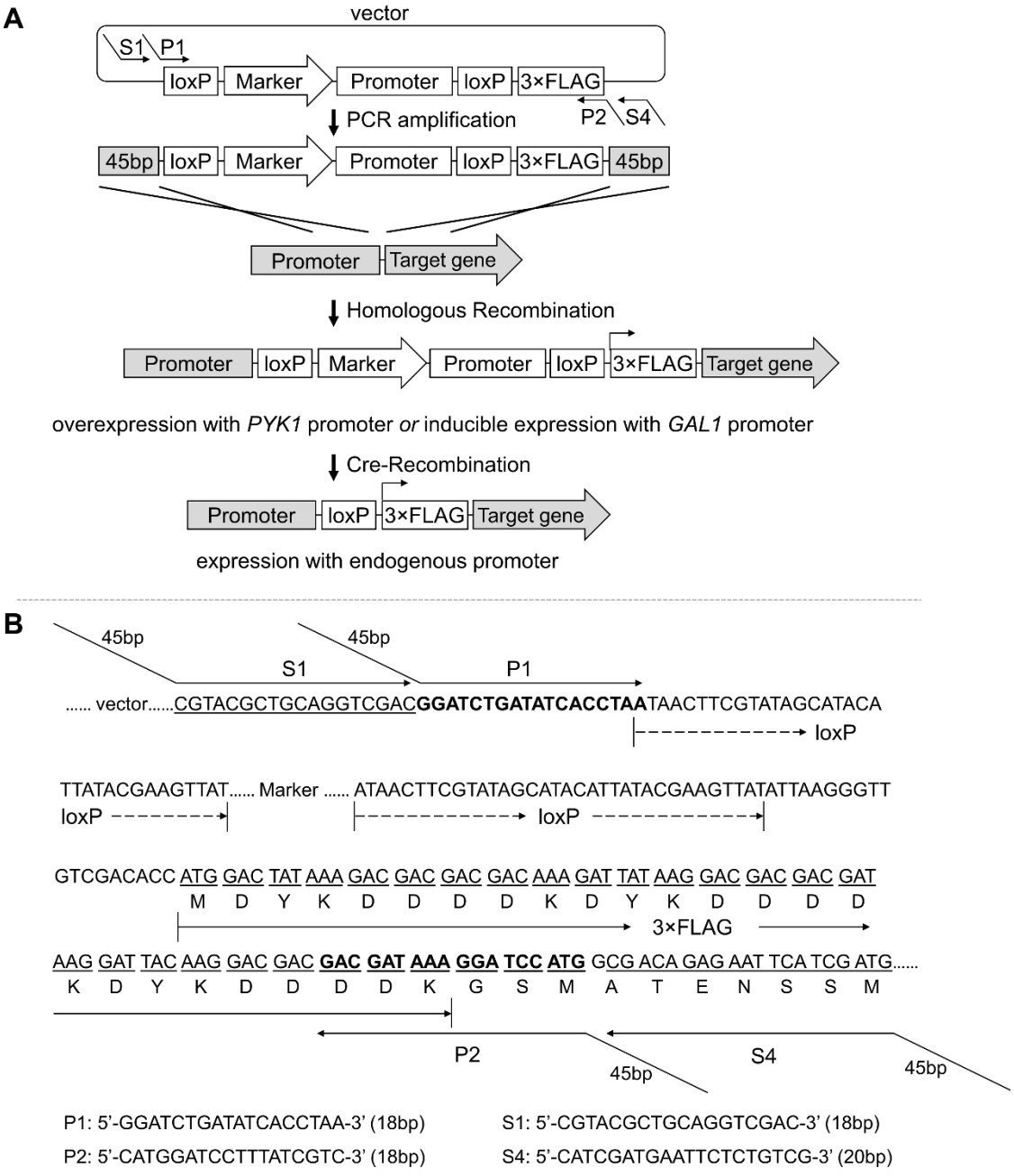


Figure 3.3 N-ICE plasmid strategy and primer design for N-terminal 3×FLAG epitope tagging.

(A) The flow diagram of the PCR-based N-terminal 3×FLAG epitope tagging strategy using N-ICE plasmids. (B) Primer design for the PCR-based N-terminal 3×FLAG epitope tagging strategy using N-ICE plasmids. DNA sequences of the loxP sites, 3×FLAG epitope, P1/P2, and S1/S4 primer sequences are presented. The selection marker (*KanMX*, *HphMX*, and *NatMX*) cassette and heterologous promoter are represented as "marker".

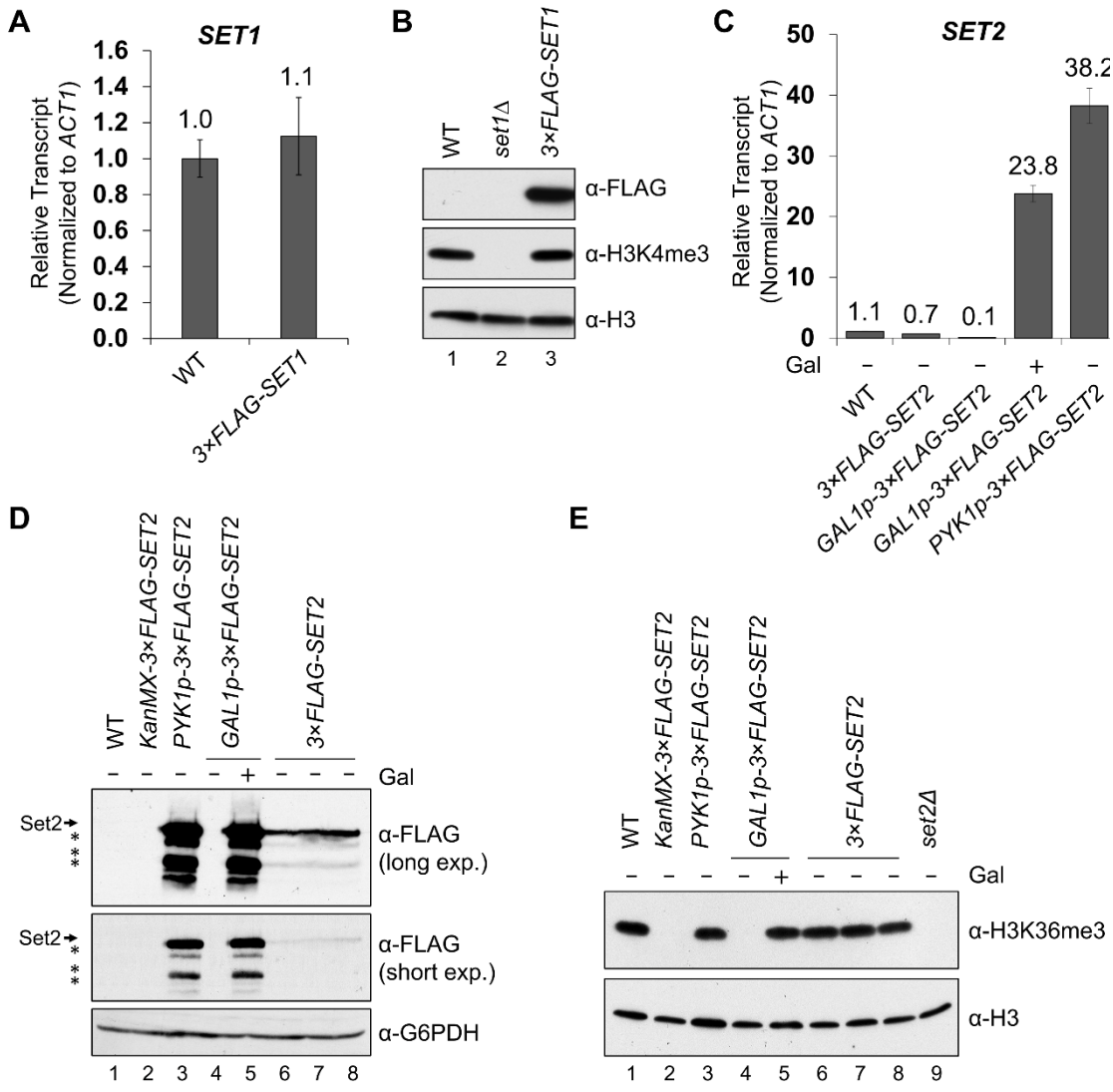


Figure 3.4 N-terminal 3×FLAG epitope tagging of Set1 and Set2 using first generation N-ICE plasmids.

(A) The relative transcript level of *SET1* is determined by qRT-PCR analysis. Expression analysis is relative to wild-type (WT) cells using Actin (*ACT1*) as an internal control to normalize transcript levels. Data was analyzed from three biological repeats (each biological repeat has three technical replicates). Error bars represent the standard error of the mean and values above bars indicate fold change relative to untagged WT. The unpaired *t*-test was performed and *p*-value ≤ 0.05 is considered significant: WT vs. *3×FLAG-SET1*: $p = 0.4191$, n=3. (B) Western blot analysis of whole cell extracts from N-terminal 3×FLAG tagged *SET1* strain. Lane 3 represents the sample from 3×FLAG tagged *SET1* strain after Cre recombination. Immunoblots using α-FLAG M2 monoclonal antibodies show that 3×FLAG-Set1 is expressed from the endogenous *SET1* promoter. WT and *set1*Δ are used as negative control. Western blot analysis of histone H3K4 trimethylation levels using H3K4

trimethyl-specific antibodies. Histone H3 is used as a loading control. (C) Relative transcript levels of *SET2* expressed from different promoters and growth conditions are determined by qRT-PCR analysis. Expression analysis is performed as previously described. Cells were grown up to O.D. 0.6 in YPD medium or YP medium with 1% galactose for 5 hours. Data was analyzed from three biological repeats (each biological repeat has three technical replicates). Error bars represent the standard error of the mean and values above bars indicate fold change relative to untagged WT. The unpaired *t*-test was performed and *p*-value ≤ 0.05 is considered significant: WT vs. *3×FLAG-SET2*: $p = 0.1812$, n = 3; *3×FLAG-SET2* vs. *GAL1p-3×FLAG-SET2* (Glu), $p < 0.0001$, n = 3; *3×FLAG-SET2* vs. *GAL1p-3×FLAG-SET2* (Gal): $p < 0.0001$, n = 3; *3×FLAG-SET2* vs. *PYK1p-3×FLAG-SET2*: $p < 0.0001$, n = 3. (D) Western blot analysis of whole cell extracts from the indicated yeast strains. Immunoblots using α-FLAG M2 monoclonal antibodies show that the 3×FLAG-Set2 protein is expressed from the endogenous *SET2* promoter and overexpressed by the integrated *PYK1* and *GAL1* promoters. Relative protein levels were quantified (figure 3.7A). G6PDH is used as a loading control. Lane 6, 7 and 8 represent samples were originated from the strain transformed with cassettes amplified from pK3F, pK3F-*PYK1*, and pK3F-*GAL1* respectively. Asterisks indicate likely protein degradation products of Set2. (E) Western blot analysis of histone H3K36 trimethylation levels using H3K36 trimethyl-specific antibodies. Lanes 1-8 are loaded in the same order as figure 3.3D. Histone H3 was used as a loading control.

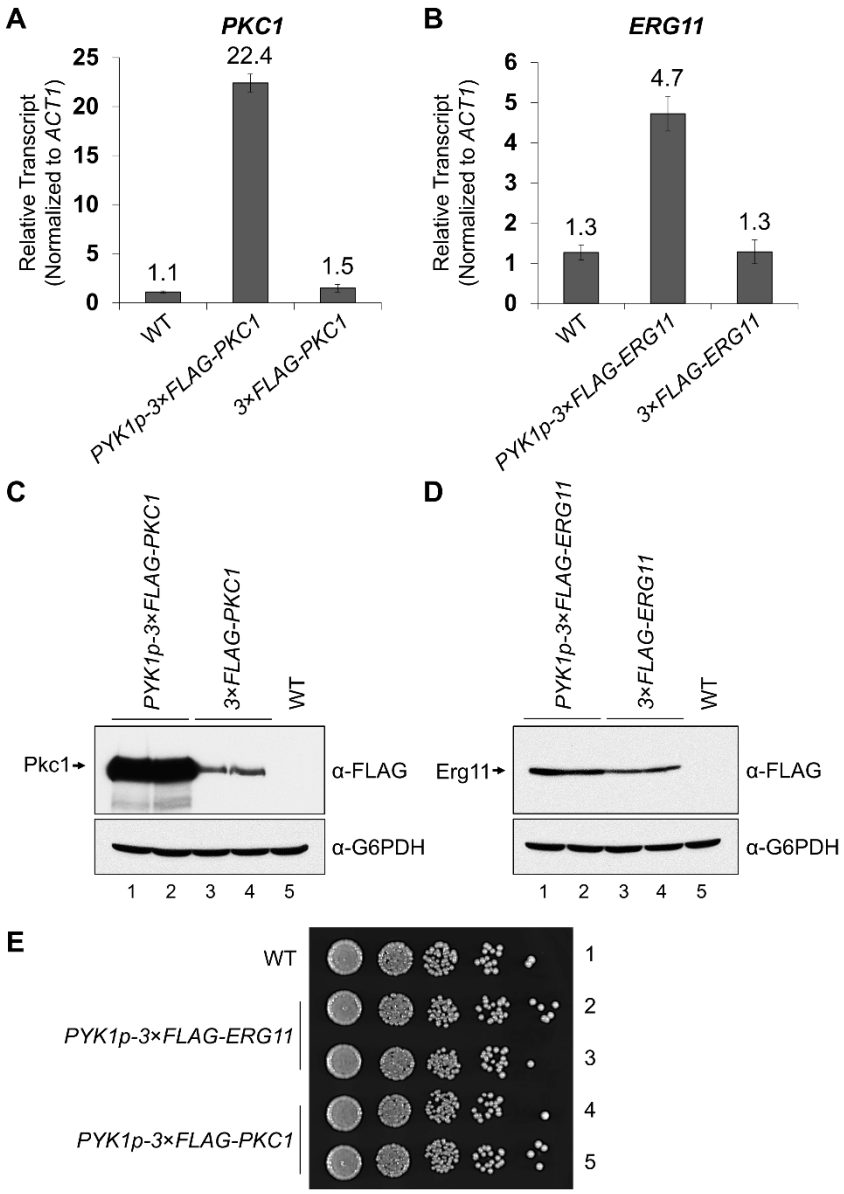


Figure 3.5 N-terminal 3×FLAG epitope tagging of essential genes, *PKC1* and *ERG11* using first generation N-ICE plasmids.

(A and B) The relative transcript levels of *PKC1* and *ERG11* are determined by qRT-PCR analysis. Data was analyzed from three biological repeats (each biological repeat has three technical replicates). Error bars represent the standard error of the mean and values above bars indicate fold change relative to untagged WT. The unpaired *t*-test was performed and *p*-value ≤ 0.05 is considered significant: WT vs. *3×FLAG-PKC1*: $p = 0.1736$, n = 3; *3×FLAG-PKC1* vs. *PYK1p-3×FLAG-PKC1*: $p < 0.001$, n = 3; WT vs. *3×FLAG-ERG11*: $p = 0.9453$, n = 3; *3×FLAG-ERG11* vs. *PYK1p-3×FLAG- ERG11*: $p < 0.001$, n =3. (C and D) Western blot analysis of whole cell extracts from the

indicated yeast strains. Immunoblots using α-FLAG M2 monoclonal antibodies show the protein expression levels of 3×FLAG-Pkc1 or 3×FLAG-Erg11 when overexpressed from the *PYK1* promoter or expressed from their endogenous promoters. Lanes 3-4 from C and D represent samples from yeast strains after Cre recombination. Relative protein levels were quantified (Figure 3.7 B and C). G6PDH is used as a loading control. The two lanes of one indicated strain represent two independent colonies. (E) Serial dilution spot assays on YPD plates of wild-type (BY4741) or yeast strains overexpressing 3×FLAG-Erg11 or 3×FLAG-Pkc1 from a *PYK*1 promoter.

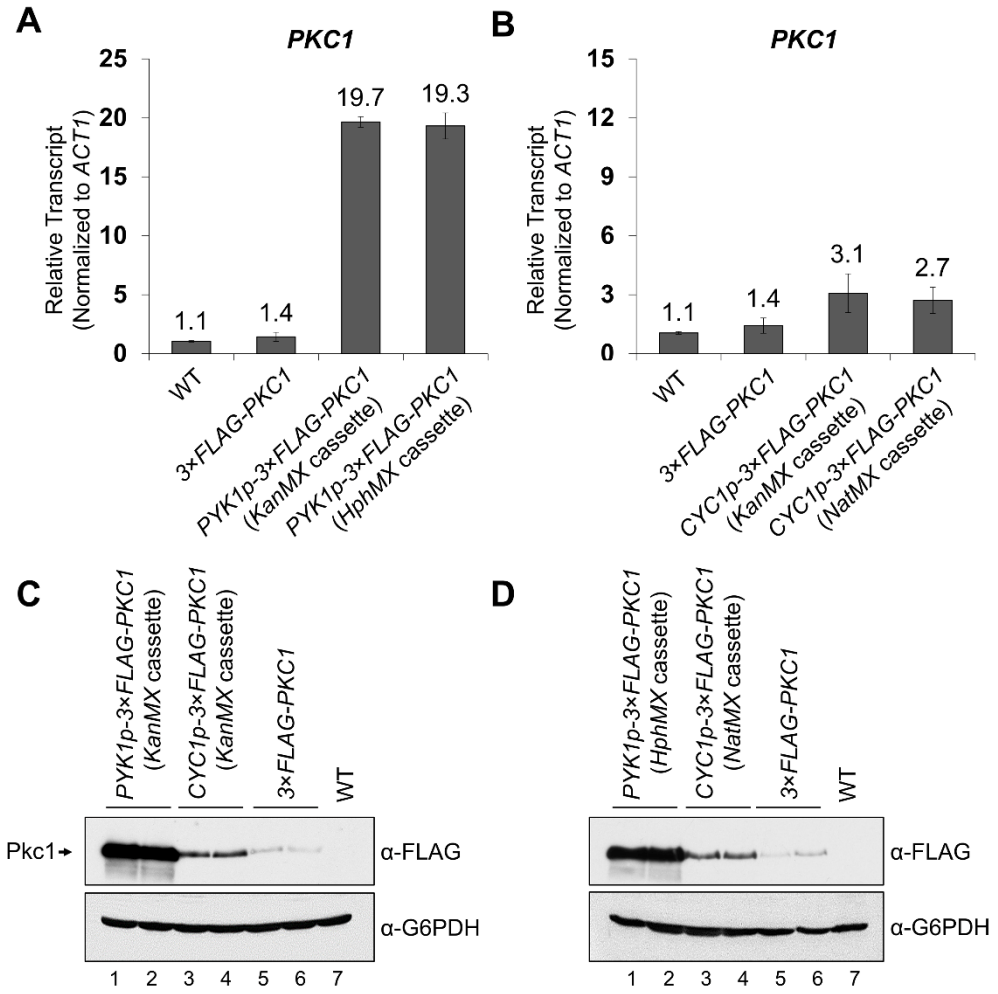


Figure 3.6 N-terminal 3×FLAG epitope tagging of essential genes *PKC1* using second generation N-ICE plasmids.

(A and B) The relative transcript levels of *PKC1* are shown as determined by qRT-PCR analysis. Data was analyzed from three biological repeats (each biological repeat has three technical replicates). Error bars represent the standard error of the mean and values above bars indicate fold change relative to untagged WT. The unpaired *t*-test was performed and *p*-value ≤ 0.05 is considered significant: WT vs. *3×FLAG-PKC1*: $p = 0.1400$, n = 3; *3×FLAG-PKC1* vs. *PYK1p-3×FLAG-PKC1(KanMX)*, $p < 0.001$, n = 3; *3×FLAG-PKC1* vs. *PYK1p-3×FLAG- PKC1(HygMX)*, $p < 0.001$, n = 3; *3×FLAG-PKC1* vs. *CYC1p-3×FLAG-PKC1 (KanMX):* $p = 0.0595$, n = 3; *3×FLAG-PKC1* vs. *CYC1p-3×FLAG-PKC1(NatMX):* $p = 0.0494$, n = 3. (C and D) Western blot analysis of whole cell extracts from the indicated yeast strains. Lanes 5-6 from C and D represent samples from yeast strains after Cre recombination. Immunoblots using α-FLAG M2 monoclonal antibodies show the protein expression levels of 3×FLAG-Pkc1 when expressed from a *PYK1* promoter, a *CYC1* promoter or the endogenous promoter. Relative protein levels were quantified (see Figure 3.7 D and E). G6PDH is used as a loading control.

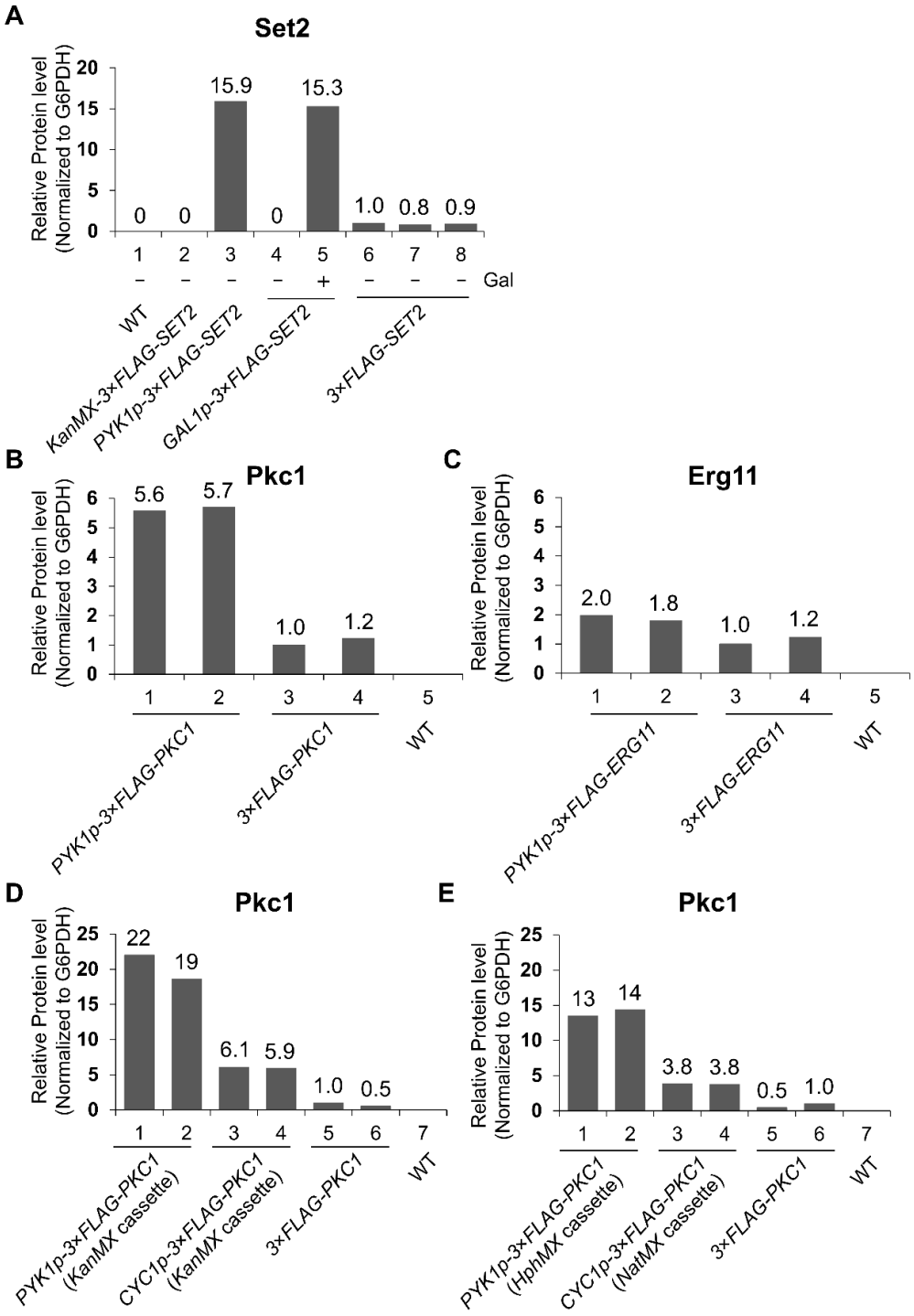


Figure 3.7 Relative quantification of Western blot signal intensity of whole cell extracts from the indicated yeast strains.

Densitometry quantification was performed with ImageJ. (A) Quantification of Western blot of whole cell extracts from the N-terminal 3×FLAG tagged *SET2* strains

(see figure 3.4D). (B) Quantification of Western blot of whole cell extracts from the N-terminal 3×FLAG tagged *PKC1* strains (see Figure 3.5C). (C) Quantification of Western blot of whole cell extracts from the N-terminal 3×FLAG tagged *ERG11* strains (see figure 3.5D). (D and E) Quantification of Western blot of whole cell extracts from the N-terminal 3×FLAG tagged *PKC1* strains using S1/S4 primer sets (see figure 3.6C and D). The protein levels were relative to the endogenous 3×FLAG tagged *PKC1* strains.

CHAPTER 4. FUTURE DIRECTIONS

4.1 Characterizing how Set1 interacts with RNAPII.

In chapter 2, we determined that Set1 interacts Ser5-P and/or Ser7-P CTD of RNAPII, and Set1-RNAPII interaction is independent of Paf1 complex and most Set1-associated subunits. However, we have not rule out the function of Swd2 in Set1-RNAPII interaction and whether combinations of complex subunits contribute to the Set1-RNAPII interaction. To determine whether Swd2 is required for Set1-RNAPII interaction, immunoprecipitation assay of Set1 with RNAII in deletion strain of *SWD2* and overexpression C-terminus of Sen1 (164). To test whether Set1 is sufficient for Set1-RNAPII interaction, *in vitro* binding assay of full length Set1 and CTD Ser5 phosphorylated peptides will be performed.

We also determined the SRIM of Set1 and requirement of SRIM for Set1-RNAPII interaction, however, the SRIM itself is not sufficient for the interaction. To define the boundary of sufficient domain, immunoprecipitation assay of more Set1 truncations and *in vitro* binding assay of full length and truncations of Set1 and CTD phosphorylated peptides will be performed.

4.2 Characterizing the dependency of RNAII in Set1 recruitment and Histone H3K4 methylation

Our set1 SRIM mutants show a great reduction of Set1-chromatin association at 5’ without affecting set1 mutant localization at promoter and 3’ of actively transcribed genes, and set1 SRIM mutants fail to be recruited by RNAPII at induced *PCK1* gene. To test whether the occupancy of SRIM mutants has genome-wide defects, ChIP-seq assay of full length Set1 and set1 SRIM mutants will be performed. To test whether set1 SRIM mutants have broader recruitment defects, other environmental stress conditions, such as heat shock and osmolality stress, can be also applied.

Our data also suggest that the deposition of H3K4 trimethylation by Set1 at transcription start site is through a RNAII recruitment manner at *PCK1* gene. To test if it is genome-wide defect, ChIP-seq assay of H3K4me3 in full length Set1 and set1 YFD→AAA mutant under different inducible conditions will be performed.

We have not identified H3K4me2 defects in set1 SRIM mutants at *PYK1* and *PMA1* genes, however, it is possible that there are defects in other genes. To test this, ChIP-seq assay of H3K4me2 in full length Set1 and set1 YFD→AAA mutant will be performed.

Our combined data suggest that Set1 can also be recruited to chromatin and methylate H3K4 through other mechanisms independent of RNAII recruitment. The RNAPII-independent Set1 recruitment could be through Set1-asscoiated subunits or other unknown mechanisms. To test these, the Set1 recruitment and H3K4 methylation in full-

length Set1 and set1 SRIM mutants will be analyzed in double deletion strains of *SET1* and each subunit.

4.3 Characterizing the biological function of RNAII recruitment of Set1

In chapter 2, our growth assay shows set1 SRIM mutants have growth defects on 6-AU plates suggesting there are potential transcriptional defects in set1 SRIM mutants. Future experiments, such as nuclear run-on assay, would be needed to show transcriptional elongation is reduced.

Other environmental stress conditions will be tested for full-length Set1 and set1 SRIM mutants. If we found a defect, gene expressions and ChIP assay will be applied to determine whether RNAPII recruitment of Set1 is required for gene expression.

4.4 N-ICE plasmid system is an expandable and valuable toolbox for yeast community.

In Chapter 3, we have generated 14 plasmids for the N-ICE plasmid system to generate N-terminal 3×FLAG epitope tagged essential and non-essential genes. There plasmids can be readily modified, for example, the 3×FLAG epitope tag sequence can be switched to other epitope tags using restriction enzyme of SalI and BamHI. Also, the selection markers and heterogenous promoters can also be switched to other markers and promoter sequences. Additionally, our system also provides a method for yeast genomic knock-in by switching the 3×FLAG epitope tag with your interested sequences. Overall,

we have provided a new and valuable N-ICE plasmid system for the yeast community and expanded the toolbox for epitope tagging.

LIST OF REFERENCES

LIST OF REFERENCES

1. Waddington CH (2012) The epigenotype. 1942. *Int J Epidemiol* 41(1):10-13.
2. Jablonka E & Lamm E (2012) Commentary: The epigenotype--a dynamic network view of development. *Int J Epidemiol* 41(1):16-20.
3. Holoch D & Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* 16(2):71-84.
4. McCue AD, Nuthikattu S, Reeder SH, & Slotkin RK (2012) Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet* 8(2):e1002474.
5. Maurange C & Paro R (2002) A cellular memory module conveys epigenetic inheritance of hedgehog expression during Drosophila wing imaginal disc development. *Genes Dev* 16(20):2672-2683.
6. Choy JS & Lee TH (2012) Structural dynamics of nucleosomes at single-molecule resolution. *Trends in Biochemical Sciences* 37(10):425-435.
7. Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* 184(4139):868-871.
8. Luger K & Hansen JC (2005) Nucleosome and chromatin fiber dynamics. *Curr Opin Struc Biol* 15(2):188-196.
9. Bassett A, Cooper S, Wu C, & Travers A (2009) The folding and unfolding of eukaryotic chromatin. *Curr Opin Genet Dev* 19(2):159-165.
10. Luger K, Mader AW, Richmond RK, Sargent DF, & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* 389(6648):251-260.
11. Burgess RJ & Zhang ZG (2013) Histone chaperones in nucleosome assembly and human disease. *Nature Structural & Molecular Biology* 20(1):14-22.
12. Smith S & Stillman B (1991) Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J* 10(4):971-980.
13. Andrews AJ & Luger K (2011) Nucleosome structure(s) and stability: variations on a theme. *Annu Rev Biophys* 40:99-117.
14. Iwasaki W, *et al.* (2013) Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* 3:363-369.
15. Davey CA, Sargent DF, Luger K, Maeder AW, & Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. *Journal of Molecular Biology* 319(5):1097-1113.

16. Dorigo B, *et al.* (2004) Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* 306(5701):1571-1573.
17. Eberharter A & Becker PB (2004) ATP-dependent nucleosome remodelling: factors and functions. *J Cell Sci* 117(Pt 17):3707-3711.
18. Flaus A & Owen-Hughes T (2001) Mechanisms for ATP-dependent chromatin remodelling. *Curr Opin Genet Dev* 11(2):148-154.
19. Flaus A & Owen-Hughes T (2011) Mechanisms for ATP-dependent chromatin remodelling: the means to the end. *Febs J* 278(19):3579-3595.
20. Nair SS & Kumar R (2012) Chromatin remodeling in cancer: a gateway to regulate gene transcription. *Mol Oncol* 6(6):611-619.
21. Clapier CR & Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78:273-304.
22. Denslow SA & Wade PA (2007) The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26(37):5433-5438.
23. Klein BJ, *et al.* (2014) The Histone-H3K4-Specific Demethylase KDM5B Binds to Its Substrate and Product through Distinct PHD Fingers. *Cell Rep* 6(2):325-335.
24. Phillips DM & Simson P (1962) Identification of some peptides from an arginine-rich histone and their bearing on the structure of deoxyribonucleohistone. . *Biochem. J.* (82):236-241.
25. Allfrey VG, Faulkner R, & Mirsky AE (1962) Acetylation and Methylation of Histones and their Possible Role in the Regulation of RNA Synthesis. 51(5):786-794.
26. Suganuma T & Workman JL (2011) Signals and combinatorial functions of histone modifications. *Annu Rev Biochem* 80:473-499.
27. Zentner GE & Henikoff S (2013) Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* 20(3):259-266.
28. Phillips DM (1963) The presence of acetyl groups of histones. *Biochem J* 87:258-263.
29. Li B, Carey M, & Workman JL (2007) The role of chromatin during transcription. *Cell* 128(4):707-719.
30. Vaquero A, Loyola A, & Reinberg D (2003) The constantly changing face of chromatin. *Sci Aging Knowledge Environ* 2003(14):RE4.
31. Brownell JE & Allis CD (1996) Special HATs for special occasions: Linking histone acetylation to chromatin assembly and gene activation. *Current Opinion in Genetics & Development* 6(2):176-184.
32. Brownell JE, *et al.* (1996) Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84(6):843-851.
33. Allfrey VG, FAULKNER R, & Mirsky AE (1964) Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. *Proc. Natl. Acad. Sci.* 51:786-794.
34. Fukuda H, Sano N, Muto S, & Horikoshi M (2006) Simple histone acetylation plays a complex role in the regulation of gene expression. *Brief Funct Genomic Proteomic* 5(3):190-208.

35. Kleff S, Andrulis ED, Anderson CW, & Sternglanz R (1995) Identification of a gene encoding a yeast histone H4 acetyltransferase. *J Biol Chem* 270(42):24674-24677.
36. Brownell JE & Allis CD (1995) An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proc Natl Acad Sci U S A* 92(14):6364-6368.
37. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128(4):693-705.
38. Liu CL, *et al.* (2005) Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol* 3(10):e328.
39. Hong L, Schroth GP, Matthews HR, Yau P, & Bradbury EM (1993) Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem* 268(1):305-314.
40. Mujtaba S, Zeng L, & Zhou MM (2007) Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* 26(37):5521-5527.
41. Zeng L & Zhou MM (2002) Bromodomain: an acetyl-lysine binding domain. *Febs Letters* 513(1):124-128.
42. Filippakopoulos P & Knapp S (2012) The bromodomain interaction module. *FEBS Lett* 586(17):2692-2704.
43. Imai S, Armstrong CM, Kaeberlein M, & Guarente L (2000) Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* 403(6771):795-800.
44. Yang XJ & Seto E (2008) The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Bio* 9(3):206-218.
45. Rossetto D, Avvakumov N, & Cote J (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* 7(10):1098-1108.
46. Rossetto D, Truman AW, Kron SJ, & Cote J (2010) Epigenetic modifications in double-strand break DNA damage signaling and repair. *Clin Cancer Res* 16(18):4543-4552.
47. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, & Bonner WM (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273(10):5858-5868.
48. Redon C, *et al.* (2003) Yeast histone 2A serine 129 is essential for the efficient repair of checkpoint-blind DNA damage. *Embo Rep* 4(7):678-684.
49. Xiao A, *et al.* (2009) WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. *Nature* 457(7225):57-62.
50. Wei Y, Mizzen CA, Cook RG, Gorovsky MA, & Allis CD (1998) Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. *P Natl Acad Sci USA* 95(13):7480-7484.
51. Wei Y, Yu LL, Bowen J, Gorovsky MA, & Allis CD (1999) Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* 97(1):99-109.

52. Goto H, *et al.* (1999) Identification of a novel phosphorylation site on histone H3 coupled with mitotic chromosome condensation. *Journal of Biological Chemistry* 274(36):25543-25549.
53. Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* 70:503-533.
54. Finley D & Chau V (1991) Ubiquitination. *Annu Rev Cell Biol* 7:25-69.
55. Fuchs G & Oren M (2014) Writing and reading H2B monoubiquitylation. *Biochim Biophys Acta* 1839(8):694-701.
56. Osley MA (2006) Regulation of histone H2A and H2B ubiquitylation. *Brief Funct Genomic Proteomic* 5(3):179-189.
57. Dover J, *et al.* (2002) Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *J Biol Chem* 277(32):28368-28371.
58. Fingerman IM, Du HN, & Briggs SD (2008) Controlling histone methylation via trans-histone pathways. *Epigenetics* 3(5):237-242.
59. Sun ZW & Allis CD (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 418(6893):104-108.
60. Briggs SD, *et al.* (2002) Gene silencing: trans-histone regulatory pathway in chromatin. *Nature* 418(6897):498.
61. Murray K (1964) The Occurrence of Epsilon-n-methyl Lysine in Histones. *Biochemistry-Us* (3):10-15.
62. Tessarz P, *et al.* (2014) Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* 505(7484):564-+.
63. Jenuwein T & Allis CD (2001) Translating the histone code. *Science* 293(5532):1074-1080.
64. Zhang Y & Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Gene Dev* 15(18):2343-2360.
65. Wozniak GG & Strahl BD (2014) Hitting the 'mark': Interpreting lysine methylation in the context of active transcription. *Bba-Gene Regul Mech* 1839(12):1353-1361.
66. Lee DY, Teyssier C, Strahl BD, & Stallcup MR (2005) Role of protein methylation in regulation of transcription. *Endocr Rev* 26(2):147-170.
67. Edwards CR, Dang WW, & Berger SL (2011) Histone H4 Lysine 20 of Saccharomyces cerevisiae Is Monomethylated and Functions in Subtelomeric Silencing. *Biochemistry-Us* 50(48):10473-10483.
68. Martin GM, *et al.* (2014) Set5 and Set1 cooperate to repress gene expression at telomeres and retrotransposons. *Epigenetics* 9(4):513-522.
69. Martin GM, King DA, Garcia-Nieto PE, & Morrison AJ (2014) Transcriptome profiling of Set5 and Set1 methyltransferases: Tools for visualization of gene expression. *Genom Data* 2:216-218.
70. Herz HM, Garruss A, & Shilatifard A (2013) SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem Sci* 38(12):621-639.
71. van Leeuwen F, Gafken PR, & Gottschling DE (2002) Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* 109(6):745-756.

72. Nguyen AT & Zhang Y (2011) The diverse functions of Dot1 and H3K79 methylation. *Genes Dev* 25(13):1345-1358.
73. Eissenberg JC & Shilatifard A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental Biology* 339(2):240-249.
74. Shi YG & Tsukada Y (2013) The Discovery of Histone Demethylases (vol 5, a017947, 2013). *Csh Perspect Biol* 5(12).
75. Shi Y, *et al.* (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119(7):941-953.
76. Ramakrishnan S, *et al.* (2016) Counteracting H3K4 methylation modulators Set1 and Jhd2 co-regulate chromatin dynamics and gene transcription. *Nat Commun* 7:11949.
77. Kooistra SM & Helin K (2012) Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol* 13(5):297-311.
78. Tsukada Y, *et al.* (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439(7078):811-816.
79. Greer EL & Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* 13(5):343-357.
80. Mellor J (2006) It takes a PHD to read the histone code. *Cell* 126(1):22-24.
81. Taverna SD, Li H, Ruthenburg AJ, Allis CD, & Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Structural & Molecular Biology* 14(11):1025-1040.
82. Shilatifard A (2008) Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr Opin Cell Biol* 20(3):341-348.
83. Yap KL & Zhou MM (2011) Structure and Mechanisms of Lysine Methylation Recognition by the Chromodomain in Gene Transcription. *Biochemistry-Us* 50(12):1966-1980.
84. Smallwood A, Esteve PO, Pradhan S, & Carey M (2007) Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Gene Dev* 21(10):1169-1178.
85. Gatchalian J, *et al.* (2016) Structural Insight into Recognition of Methylated Histone H3K4 by Set3. *J Mol Biol*.
86. Yu Y, Zhou H, Deng X, Wang W, & Lu H (2016) Set3 contributes to heterochromatin integrity by promoting transcription of subunits of Clr4-Rik1-Cul4 histone methyltransferase complex in fission yeast. *Sci Rep* 6:31752.
87. Kim T & Buratowski S (2009) Dimethylation of H3K4 by Set1 Recruits the Set3 Histone Deacetylase Complex to 5 ' Transcribed Regions. *Cell* 137(2):259-272.
88. Di Lorenzo A & Bedford MT (2011) Histone arginine methylation. *FEBS Lett* 585(13):2024-2031.
89. Wysocka J, Allis CD, & Coonrod S (2006) Histone arginine methylation and its dynamic regulation. *Front Biosci* 11:344-355.
90. Chang BS, Chen Y, Zhao YM, & Bruick RK (2007) JMJD6 is a histone arginine demethylase. *Science* 318(5849):444-447.
91. Di Lorenzo A & Bedford MT (2011) Histone arginine methylation. *Febs Letters* 585(13):2024-2031.
92. Santos-Rosa H, *et al.* (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* 419(6905):407-411.

93. Becker PB (2006) Gene regulation: a finger on the mark. *Nature* 442(7098):31-32.
94. Ruthenburg AJ, Allis CD, & Wysocka J (2007) Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* 25(1):15-30.
95. Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, 3rd, & Grant PA (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 433(7024):434-438.
96. Sims RJ, *et al.* (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *Journal of Biological Chemistry* 280(51):41789-41792.
97. Barski A, *et al.* (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823-837.
98. Zhang X, *et al.* (2002) Structure of the neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* 111(1):117-127.
99. Alvarez-Venegas R & Avramova Z (2002) SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. *Gene* 285(1-2):25-37.
100. Deng P, Chen QM, Hong C, & Wang CY (2015) Histone methyltransferases and demethylases: regulators in balancing osteogenic and adipogenic differentiation of mesenchymal stem cells. *Int J Oral Sci* 7(4):197-204.
101. Zhang X, *et al.* (2003) Structural basis for the product specificity of histone lysine methyltransferases. *Mol Cell* 12(1):177-185.
102. Min J, Zhang X, Cheng X, Grewal SI, & Xu RM (2002) Structure of the SET domain histone lysine methyltransferase Clr4. *Nat Struct Biol* 9(11):828-832.
103. Briggs SD, *et al.* (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Gene Dev* 15(24):3286-3295.
104. Porras-Yakushi TR, Whitelegge JP, Miranda TB, & Clarke S (2005) A novel SET domain methyltransferase modifies ribosomal protein Rpl23ab in yeast. *Journal of Biological Chemistry* 280(41):34590-34598.
105. Strahl BD, *et al.* (2002) Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol* 22(5):1298-1306.
106. Roguev A, *et al.* (2001) The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J* 20(24):7137-7148.
107. Nagy PL, Griesenbeck J, Kornberg RD, & Cleary ML (2002) A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. *P Natl Acad Sci USA* 99(1):90-94.
108. Bryk M, *et al.* (2002) Evidence that SET1, a factor required for methylation of histone H3, regulates rDNA silencing in S-cerevisiae by a sir2-independent mechanism. *Current Biology* 12(2):165-170.
109. Miller T, *et al.* (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* 98(23):12902-12907.

110. Mersman DP, Du HN, Fingerman IM, South PF, & Briggs SD (2012) Charge-based Interaction Conserved within Histone H3 Lysine 4 (H3K4) Methyltransferase Complexes Is Needed for Protein Stability, Histone Methylation, and Gene Expression. *Journal of Biological Chemistry* 287(4):2652-2665.
111. Zhang K, *et al.* (2005) The Set1 methyltransferase opposes Ipl1 aurora kinase functions in chromosome segregation. *Cell* 122(5):723-734.
112. Latham JA, Chosed RJ, Wang S, & Dent SY (2011) Chromatin signaling to kinetochores: transregulation of Dam1 methylation by histone H2B ubiquitination. *Cell* 146(5):709-719.
113. Fingerman IM, Wu CL, Wilson BD, & Briggs SD (2005) Global loss of Set1-mediated H3 Lys4 trimethylation is associated with silencing defects in Saccharomyces cerevisiae. *J Biol Chem* 280(31):28761-28765.
114. Dehe PM, *et al.* (2006) Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *Journal of Biological Chemistry* 281(46):35404-35412.
115. Kim J, *et al.* (2013) The n-SET Domain of Set1 Regulates H2B Ubiquitylation-Dependent H3K4 Methylation. *Molecular Cell* 49(6):1121-1133.
116. Takahashi YH, *et al.* (2011) Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc Natl Acad Sci U S A* 108(51):20526-20531.
117. Lee JH & Skalnik DG (2008) Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol* 28(2):609-618.
118. Tresaugues L, *et al.* (2006) Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation. *Journal of Molecular Biology* 359(5):1170-1181.
119. South PF, Fingerman IM, Mersman DP, Du HN, & Briggs SD (2010) A conserved interaction between the SDI domain of Bre2 and the Dpy-30 domain of Sdc1 is required for histone methylation and gene expression. *J Biol Chem* 285(1):595-607.
120. Chen Y, Cao F, Wan BB, Dou YL, & Lei M (2012) Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. *Cell Res* 22(3):598-602.
121. Cao F, *et al.* (2010) An Ash2L/RbBP5 Heterodimer Stimulates the MLL1 Methyltransferase Activity through Coordinated Substrate Interactions with the MLL1 SET Domain. *Plos One* 5(11).
122. Yu LH, Gaitatzes C, Neer E, & Smith TE (2000) Thirty-plus functional families from a single motif. *Protein Science* 9(12):2470-2476.
123. Patel A, Dharmarajan V, & Cosgrove MS (2008) Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem* 283(47):32158-32161.
124. Nedea E, *et al.* (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol Cell* 29(5):577-587.

125. Halbach A, *et al.* (2009) Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *EMBO J* 28(19):2959-2970.
126. Shi XB, *et al.* (2007) Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *Journal of Biological Chemistry* 282(4):2450-2455.
127. Murton BL, Chin WL, Ponting CP, & Itzhaki LS (2010) Characterising the Binding Specificities of the Subunits Associated with the KMT2/Set1 Histone Lysine Methyltransferase. *Journal of Molecular Biology* 398(4):481-488.
128. Nislow C, Ray E, & Pillus L (1997) SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol Biol Cell* 8(12):2421-2436.
129. Santos-Rosa H, Bannister AJ, Dehe PM, Geli V, & Kouzarides T (2004) Methylation of H3 lysine 4 at euchromatin promotes Sir3p association with heterochromatin. *Journal of Biological Chemistry* 279(46):47506-47512.
130. Venkatasubrahmanyam S, Hwang WW, Meneghini MD, Tong AHY, & Madhani HD (2007) Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z. *P Natl Acad Sci USA* 104(42):16609-16614.
131. Faucher D & Wellinger RJ (2010) Methylated H3K4, a Transcription-Associated Histone Modification, Is Involved in the DNA Damage Response Pathway. *Plos Genetics* 6(8).
132. Garcia P, del Dedo JE, Ayte J, & Hidalgo E (2016) Genome-wide Screening of Regulators of Catalase Expression ROLE OF A TRANSCRIPTION COMPLEX AND HISTONE AND tRNA MODIFICATION COMPLEXES ON ADAPTATION TO STRESS. *Journal of Biological Chemistry* 291(2):790-799.
133. South PF, Harmeyer KM, Serratore ND, & Briggs SD (2013) H3K4 methyltransferase Set1 is involved in maintenance of ergosterol homeostasis and resistance to Brefeldin A. *Proc Natl Acad Sci U S A* 110(11):E1016-1025.
134. Young RA (1991) Rna Polymerase-Ii. *Annual Review of Biochemistry* 60:689-715.
135. Cramer P, *et al.* (2008) Structure of eukaryotic RNA polymerases. *Annual Review of Biophysics* 37:337-352.
136. Carlson M (1997) Genetics of transcriptional regulation in yeast: connections to the RNA polymerase II CTD. *Annu Rev Cell Dev Biol* 13:1-23.
137. Nonet M, Sweetser D, & Young RA (1987) Functional Redundancy and Structural Polymorphism in the Large Subunit of Rna Polymerase-Ii. *Cell* 50(6):909-915.
138. Allison LA, Wong JKC, Fitzpatrick VD, Moyle M, & Ingles CJ (1988) The C-Terminal Domain of the Largest Subunit of Rna Polymerase-Ii of Saccharomyces-Cerevisiae, Drosophila-Melanogaster, and Mammals - a Conserved Structure with an Essential Function. *Mol Cell Biol* 8(1):321-329.
139. Nonet ML & Young RA (1989) Intragenic and Extragenic Suppressors of Mutations in the Heptapeptide Repeat Domain of Saccharomyces-Cerevisiae Rna Polymerase-Ii. *Genetics* 123(4):715-724.

140. Napolitano G, Lania L, & Majello B (2014) RNA Polymerase II CTD Modifications: How Many Tales From a Single Tail. *Journal of Cellular Physiology* 229(5):538-544.
141. Kizer KO, *et al.* (2005) A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3K36 methylation with transcript elongation. *Mol Cell Biol* 25(8):3305-3316.
142. Mayer A, *et al.* (2012) CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science* 336(6089):1723-1725.
143. Hintermair C, *et al.* (2016) Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression. *Sci Rep* 6:27401.
144. Krogan NJ, *et al.* (2002) RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol Cell Biol* 22(20):6979-6992.
145. Mayer A, *et al.* (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17(10):1272-1278.
146. Wade PA, *et al.* (1996) A novel collection of accessory factors associated with yeast RNA polymerase II. *Protein Expr Purif* 8(1):85-90.
147. Crisucci EM & Arndt KM (2011) The Roles of the Paf1 Complex and Associated Histone Modifications in Regulating Gene Expression. *Genet Res Int* 2011.
148. Robzyk K, Recht J, & Osley MA (2000) Rad6-dependent ubiquitination of histone H2B in yeast. *Science* 287(5452):501-504.
149. Penheiter KL, Washburn TM, Porter SE, Hoffman MG, & Jaehning JA (2005) A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. *Molecular Cell* 20(2):213-223.
150. Krogan NJ, *et al.* (2003) The Paf1 complex is required for histone h3 methylation by COMPASS and Dot1p: Linking transcriptional elongation to histone methylation. *Molecular Cell* 11(3):721-729.
151. Wood A, Schneider J, Dover J, Johnston M, & Shilatifard A (2003) The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *Journal of Biological Chemistry* 278(37):34739-34742.
152. Shilatifard A (2012) The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annual Review of Biochemistry, Vol 81* 81:65-95.
153. Ng HH, Robert F, Young RA, & Struhl K (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 11(3):709-719.
154. Santos-Rosa H, *et al.* (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* 419(6905):407-411.
155. Shilatifard A (2008) Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr Opin Cell Biol* 20(3):341-348.
156. Hughes CM, *et al.* (2004) Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell* 13(4):587-597.
157. Milne TA, *et al.* (2005) MLL associates specifically with a subset of transcriptionally active target genes. *Proc Natl Acad Sci U S A* 102(41):14765-14770.

158. Hampsey M & Reinberg D (2003) Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* 113(4):429-432.
159. Orphanides G & Reinberg D (2002) A unified theory of gene expression. *Cell* 108(4):439-451.
160. Cole MD & Cowling VH (2008) Transcription-independent functions of MYC: regulation of translation and DNA replication. *Nat Rev Mol Cell Biol* 9(10):810-815.
161. Ng HH, Robert F, Young RA, & Struhl K (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 11(3):709-719.
162. Wood A, Schneider J, Dover J, Johnston M, & Shilatifard A (2003) The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem* 278(37):34739-34742.
163. Jones JC, *et al.* (2004) C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. *J Biol Chem* 279(24):24957-24964.
164. Nedea E, *et al.* (2008) The glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Molecular Cell* 29(5):577-587.
165. Du HN & Briggs SD (2010) A Nucleosome Surface Formed by Histone H4, H2A, and H3 Residues Is Needed for Proper Histone H3 Lys(36) Methylation, Histone Acetylation, and Repression of Cryptic Transcription. *Journal of Biological Chemistry* 285(15):11704-11713.
166. Gelbart ME, Rechsteiner T, Richmond TJ, & Tsukiyama T (2001) Interactions of Isw2 chromatin remodeling complex with nucleosomal arrays: Analyses using recombinant yeast histones and immobilized templates. *Mol Cell Biol* 21(6):2098-2106.
167. Winzeler EA, *et al.* (1999) Functional characterization of the S-cerevisiae genome by gene deletion and parallel analysis. *Science* 285(5429):901-906.
168. Goldstein AL & McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast* 15(14):1541-1553.
169. Briggs SD, *et al.* (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev* 15(24):3286-3295.
170. Harmeyer KM, South PF, Bishop B, Ogas J, & Briggs SD (2015) Immediate chromatin immunoprecipitation and on-bead quantitative PCR analysis: a versatile and rapid ChIP procedure. *Nucleic Acids Res* 43(6):e38.
171. Wach A, Brachat A, Pohlmann R, & Philippsen P (1994) New Heterologous Modules for Classical or Pcr-Based Gene Disruptions in Saccharomyces-Cerevisiae. *Yeast* 10(13):1793-1808.
172. Baudin A, Ozierkalogeropoulos O, Denouel A, Lacroute F, & Cullin C (1993) A Simple and Efficient Method for Direct Gene Deletion in Saccharomyces-Cerevisiae. *Nucleic Acids Research* 21(14):3329-3330.

173. Johnston M, Riles L, & Hegemann JH (2002) Gene disruption. *Method Enzymol* 350:290-315.
174. Janke C, *et al.* (2004) A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* 21(11):947-962.
175. Knop M, *et al.* (1999) Epitope tagging of yeast genes using a PCR-based strategy: More tags and improved practical routines. *Yeast* 15(10b):963-972.
176. De Antoni A & Gallwitz D (2000) A novel multi-purpose cassette for repeated integrative epitope tagging of genes in Saccharomyces cerevisiae. *Gene* 246(1-2):179-185.
177. Li SJ & Hochstrasser M (2000) The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Mol Cell Biol* 20(7):2367-2377.
178. Roguev A, *et al.* (2001) The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *Embo J* 20(24):7137-7148.
179. Gauss R, Trautwein M, Sommer T, & Spang A (2005) New modules for the repeated internal and N-terminal epitope tagging of genes in Saccharomyces cerevisiae (vol 22, pg 1, 2005). *Yeast* 22(12):1011-1011.
180. Guldener U, Heck S, Fiedler T, Beinhauer J, & Hegemann JH (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Research* 24(13):2519-2524.
181. Longtine MS, *et al.* (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* 14(10):953-961.
182. Vorvis C, Markus SM, & Lee WL (2008) Photoactivatable GFP tagging cassettes for protein-tracking studies in the budding yeast Saccharomyces cerevisiae. *Yeast* 25(9):651-659.
183. van Werven FJ & Timmers HTM (2006) The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. *Nucleic Acids Research* 34(4).
184. Zhang L, Hernan R, & Brizzard B (2001) Multiple tandem epitope tagging for enhanced detection of protein expressed in mammalian cells. *Mol Biotechnol* 19(3):313-321.
185. Moqtaderi Z & Struhl K (2008) Expanding the repertoire of plasmids for PCR-mediated epitope tagging in yeast. *Yeast* 25(4):287-292.
186. Noguchi C, Garabedian MV, Malik M, & Noguchi E (2008) A vector system for genomic FLAG epitope-tagging in Schizosaccharomyces pombe. *Biotechnol J* 3(9-10):1280-1285.
187. Khmelinskii A, Meurer M, Duishoev N, Delhomme N, & Knop M (2011) Seamless gene tagging by endonuclease-driven homologous recombination. *PLoS One* 6(8):e23794.
188. Mumberg D, Muller R, & Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156(1):119-122.

189. Espinet C, de la Torre MA, Aldea M, & Herrero E (1995) An efficient method to isolate yeast genes causing overexpression-mediated growth arrest. *Yeast* 11(1):25-32.
190. Strahl BD, *et al.* (2002) Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol* 22(5):1298-1306.
191. Du HN, Fingerman IM, & Briggs SD (2008) Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4. *Gene Dev* 22(20):2786-2798.
192. Watanabe M, Chen CY, & Levin DE (1994) Saccharomyces-Cerevisiae Pkc1 Encodes a Protein-Kinase-C (Pkc) Homolog with a Substrate-Specificity Similar to That of Mammalian Pkc. *Journal of Biological Chemistry* 269(24):16829-16836.
193. Cabello-Hurtado F, *et al.* (1999) Optimized expression and catalytic properties of a wheat obtusifoliol 14 alpha-demethylase (CYP51) expressed in yeast - Complementation of erg11 Delta yeast mutants by plant CYP51. *Eur J Biochem* 262(2):435-446.
194. Sherman F (2002) Getting started with yeast. *Method Enzymol* 350:3-41.
195. Gietz D, Stjean A, Woods RA, & Schiestl RH (1992) Improved Method for High-Efficiency Transformation of Intact Yeast-Cells. *Nucleic Acids Research* 20(6):1425-1425.
196. Carrozza MJ, *et al.* (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123(4):581-592.
197. Keogh MC, *et al.* (2005) Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123(4):593-605.
198. Henikoff S & Ahmad K (2005) Assembly of variant histones into chromatin. *Annu Rev Cell Dev Bi* 21:133-153.
199. Santisteban MS, Kalashnikova T, & Smith MM (2000) Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes. *Cell* 103(3):411-422.
200. Mizuguchi G, *et al.* (2004) ATP-Driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 303(5656):343-348.

APPENDICES

Appendix A Unpublished Data

Declaration of collaborative work

In this study, all the experiments and analysis were performed by Yueping Zhang.

The phenotypic study of epigenetics factors Rph1 in *Saccharomyces cerevisiae*

Rph1 is a JmjC domain-containing histone demethylase, targeting H3K36 tri- and dimethylation, which is deposited by Set2 (190). Set2 interacts with ser2-P and Ser5-P CTD of RNAPII during transcription elongation, and H3K36 trimethylation recruits Rpd3 histone deacetylase (HDAC) which plays an important role in preventing cryptic transcription (196, 197).

To determine the biological function of Rph1 in yeast growth under different environment stresses, the deletion of *RPH1* is generated in wild-type BY4741 and BY4742. A growth assay was performed by spotting a 5-fold serial dilution of the indicated log phase yeast strains on indicated plates (Figure A1A). The result showed there is no growth difference between *rph1*Δ and wild-type cells in the tested conditions, including changing glucose to galactose, sucrose, ethanol, and acetate, and treatment of anti-fungal drug ketoconazole. Because Rph1 removes the H3K36 di- and trimethylation deposited by Set2, it is possible that the double deletion of *RPH1* and *SET2* would lead to a more severe phenotype. A growth assay was performed by spotting a 5-fold serial dilution of the indicated log phase yeast strains on indicated plates (Figure A1B). The

result showed there is a very slight growth defect of *rph1*Δ *set2*Δ strain on acetate plate and high salt (0.6M NaCl) plate compared to single deletion strains and wild-type strain.

The deletion of *RPH1* or double deletion of *RPH1* and *SET2* does not show a severe growth defect under the tested conditions. It is possible that Rph1 and Set2 are not functions in these related pathway. Different conditions, such as cryptic transcription, need to be tested in the future.

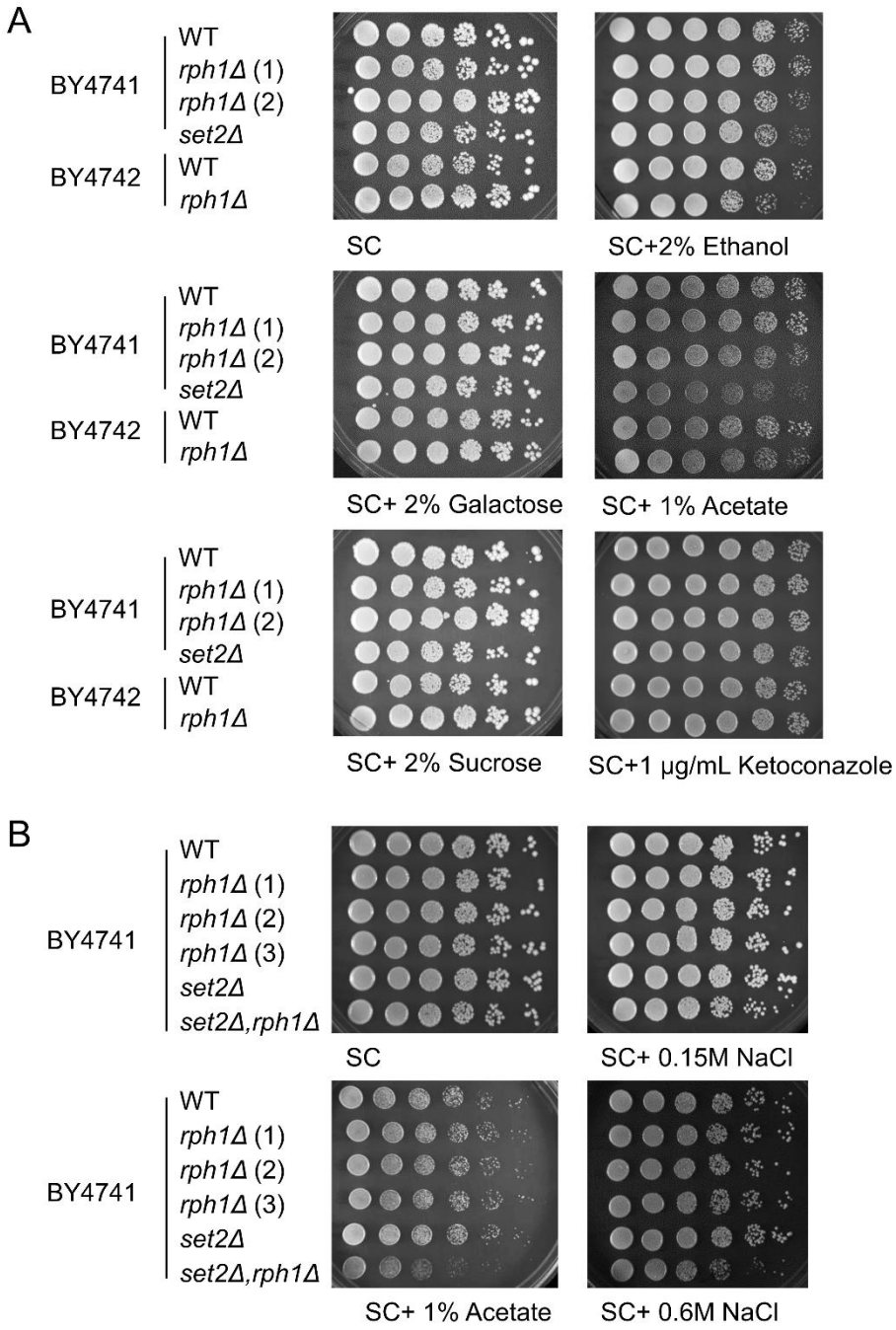


Figure A1 The growth assay of deletion of *RPH1* on indicated conditions

(A) Serial dilution spot assays wild-type (BY4741 or BY4742) or indicated strains on SC plate or SC with indicated carbon source or chemical. (B) Serial dilution spot assays wild-type (BY4741) or indicated strains on SC plate or SC with acetate or various concentrations of NaCl.

The phenotypic study of epigenetics factors Htz1 in *Saccharomyces cerevisiae*

Histone Htz1(H2A.Z) is an H2A variant, functions in transcription activation and anti-silencing (130, 198, 199). Htz1 is deposited on the nucleosome by SWR1 complex replacing H2A-H2B dimer with Htz1-H2B dimer (198, 200). In Saccharomyces cerevisiae, Htz1 is nonessential gene, and together with H3K4 trimethylation to antagonize the spreading of Sir proteins (HDACs) from the silencing region (130).

To determine the biological function of Htz1 in yeast, we generated *htz1*Δ in wild-type FY2609 and BY4741 strain. A growth assay was performed by spotting a 5-fold serial dilution of the indicated log phase yeast strains on indicated plates (Figure A2A). The results showed that the deletion of *HTZ1* results in a slow growth phenotype under acetate condition or at 24°C both in FY2609 and BY4741 background. Acetate is non-fermentative carbon source for yeast. Because Htz1 and Set1 coincide at transcriptional start site, it is possible that the double deletion of *HTZ1* and *SET1* results in a more severe growth phenotype on acetate condition. A growth assay showed double deletion of *HTZ1* and *SET1* is slightly slower than *htz1*Δ (Figure A2B).

The deletion of *HTZ1* exhibits a slightly slow growth defect when yeast are growing in either 24°C or 37°C which is not yeast favorite condition and in Acetate condition. The double deletion of *HTZ1* and *SET1* only showed a slightly slower than the single deletion. In the future, Microarray or RNA seq method will be applied to study what genes are responsible for this slow growth.

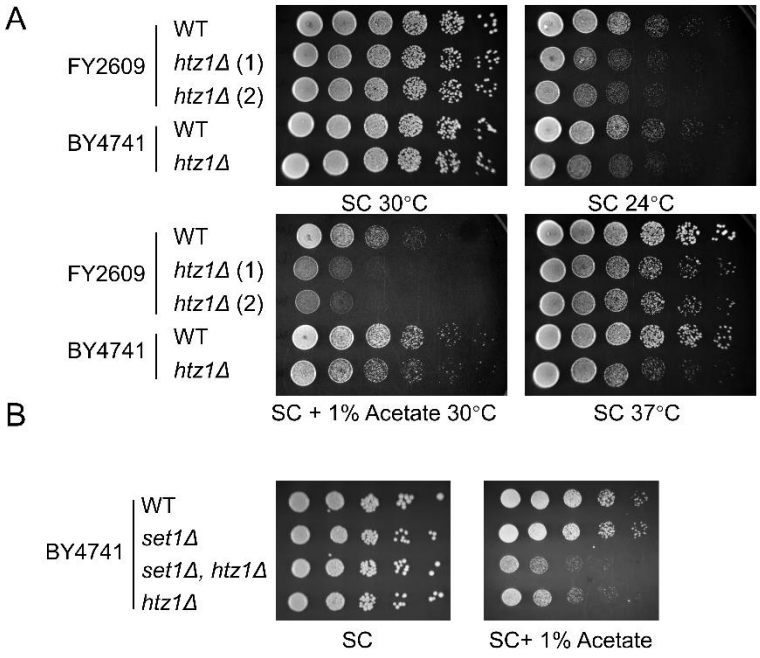


Figure A2 The deletion of *HTZ1* exhibits a temperature sensitive phenotype and slight growth defect on acetate condition

(A) Serial dilution spot assays wild-type (FY2609 or BY4741) or indicated strains on SC plate with different incubation temperatures or SC with acetate as carbon source. (B) Serial dilution spot assays wild-type (BY4741) or indicated strains on SC plate or SC with acetate as carbon source at 30°C.

Appendix B Primers and Plasmids

Table A1 Primers

Set1 260	CCGACTAGTGTTTTGGCCAATTTTATTTA
Set1 261	CGGAGATCTGTCGACACCATGGACTATAAAG
Set1 262	CGGAAGCTTATACAGTTTAGTATTTATCTTTACC
Set1 263	GGGCTCGAGTCAGTTTTCTGCTTCTTTCTCTTTAGCCTTCTTC
Set1 264	GGGCTCGAGTCATATTTCACATGAAGGGGGAGGC
Set1 265	GGGCTCGAGTCACTTTTTAAAATAGTTTTTTATTGATATGTCTTGG
Set1 266	CGGAGATCTTATTTTGATCCCATAAAAGG
set1 276	ATCTTTATAAGAGGTCTCTGCGTTTAGAGAGACACCATGGACTATAAAGACGACGACGAC
set1 277	GTCGTCGTCGTCTTTATAGTCCATGGTGTCTCTCTAAACGCAGAGACCTCTTATAAAGAT
set1 278	ATGACATCTTCATCTTCTGAAGAAGAGGAATGACGATGATACATTGATTTGTTTGGAGCT
set1 279	AGCTCCAAACAAATCAATGTATCATCGTCATTCCTCTTCTTCAGAAGATGAAGATGTCAT
set1 282	AGGACGACGACGATAAAGGATCCTGGGAAGAAGCTCCAGATAAGAAATTCAAGAG
set1 283	CTCTTGAATTTCTTATCTGGAGCTTCTTCCCAGGATCCTTTATCGTCGTCGTCCT
set1 284	TACAAGGACGACGACGATAAAGGATCCTGGTATAACCATAATGATGGTACAAGGCGACGC
set1 285	GCGTCGCCTTGTACCATCATTATGGTTATACCAGGATCCTTTATCGTCGTCGTCCTTGTA
set1 286	TACAAGGACGACGACGATAAAGGATCCTGGTCTCAATCACGGTATTCAAATAGCAATGTT
set1 287	AACATTGCTATTTGAATACCGTGATTGAGACCAGGATCCTTTATCGTCGTCGTCCTTGTA
set1 288	TACAAGGACGACGACGATAAAGGATCCTGGTATTTTGATCCCATAAAAGGCGAGTTCTTC
set1 289	GAAGAACTCGCCTTTTATGGGATCAAAATACCAGGATCCTTTATCGTCGTCGTCCTTGTA
set1 290	CAAGGACGACGACGATAAAGGATCCTGGTCAAACGGTATCATGACTAGCGGAAAAGTGGC
set1 291	GCCACTTTTCCGCTAGTCATGATACCGTTTGACCAGGATCCTTTATCGTCGTCGTCCTTG
set1 294	GACGATAAAGGATCCTGGTCAAATTTACAGGGCAAAGATATTACCTTG

Table A1 continued

set1 295	CAAGGTAATATCTTTGCCCTGTAAATTTGACCAGGATCCTTTATCGTC
set1 298	AAGAAGGCTAAAGAGAAAGAAGCAGAAAACTGACGATGATACATTGATTTGTTTGGAGCT
set1 299	AGCTCCAAACAAATCAATGTATCATCGTCAGTTTTCTGCTTCTTTCTCTTTAGCCTTCTT
set1 300	CCACAGAAAATGCAGGAACTCACTAATTTTAAACAAGCATTCCATTTTG
set1 301	CAAAATGGAATGCTTGTTTAAAATTAGTGAGTTCCTGCATTTTCTGTGG
set1 318	AAGAAATCTGACATCAGTAGTCGGAGGGGCTATTTTGATCCCATAAAAGGCGAGTTCTTC
set1 319	GAAGAACTCGCCTTTTATGGGATCAAAATAGCCCCTCCGACTACTGATGTCAGATTTCTT
set1 320	ACAGAATCTGCTCTGCTTTTGCAACAAGCTGCACCTTCAGTTTTGAGATACAACACAGAT
set1 321	ATCTGTGTTGTATCTCAAAACTGAAGGTGCAGCTTGTTGCAAAAGCAGAGCAGATTCTGT
set1 322	TTGAGATACAACACAGATAATTTGAAGTCTCCCATAAAAGGCGAGTTCTTCAATAAGGAT
set1 323	ATCCTTATTGAAGAACTCGCCTTTTATGGGAGACTTCAAATTATCTGTGTTGTATCTCAA
set1 324	GATACAACACAGATAATTTGAAGTCTGCGGCCGCATATTTTGATCCCATAAAAGGCGAGT
set1 325	ACTCGCCTTTTATGGGATCAAAATATGCGGCCGCAGACTTCAAATTATCTGTGTTGTATC
set1 326	CACAGATAATTTGAAGTCTAAGTTTCATGCGGCCGCACCCATAAAAGGCGAGTTCTTCAA
set1 327	TTGAAGAACTCGCCTTTTATGGGTGCGGCCGCATGAAACTTAGACTTCAAATTATCTGTG
set1 330	CACAGAATCTGCTCTGCTTTTGCAACAAGAACCACCTTCAGTTTTGAGATACAACACAGA
set1 331	TCTGTGTTGTATCTCAAAACTGAAGGTGGTTCTTGTTGCAAAAGCAGAGCAGATTCTGTG
set1 332	CTGCTCTGCTTTTGCAACAAAGACCACCTGCAGTTTTGAGATACAACACAGATAATTTGA
set1 333	TCAAATTATCTGTGTTGTATCTCAAAACTGCAGGTGGTCTTTGTTGCAAAAGCAGAGCAG
set1 336	ACACAGATAATTTGAAGTCTAAGTTTCATGCTTTTGATCCCATAAAAGGCGAGTTCTTCA
set1 337	TGAAGAACTCGCCTTTTATGGGATCAAAAGCATGAAACTTAGACTTCAAATTATCTGTGT

Table A1 continued

set1 338	ACAACACAGATAATTTGAAGTCTAAGTTTGCTTATTTTGATC CCATAAAAGGCGAGTTCT
set1 339	AGAACTCGCCTTTTATGGGATCAAAATAAGCAAACTTAGAC TTCAAATTATCTGTGTTGT
NLS-GST F	GGGGGTCTAGATGCCAAAGAAGAAAAGAAAGGTCGCTTCC CCTATACTAGGTTATTGGA
GST-TEV R	CCCCGGATCCCTGGAAGTACAGGTTTTCATCCGATTTTGGA GGATGGTC
CTD F BamHI	CCCCGGATCCTTTTCTCCAACTTCCCCAAC
CTD 1R XhoI	GGGGCTCGAGAGCTTAGAAGTTGGACGGACG
S5A F	GCCCGGATCCTTTTCTCCAACTGCC
S2A F	GCCCGGATCCTTTGCACCAACTTC
CTD 2R	GCGGCTCGAGAAGCTTATCTGGAATTTTC
H3 3F	GGCACTCGAGGGAAGCATGGCTAGGAC
H3 R	GTCGAGATCTCTAGACCCGGGATAAC
TTEF R	GGGTCTAGAGATCTCGACACTG
PTEF F	GATCCGCGGCCGCTACCGTTCGTATAGCATACATTATACGA AGTTATGTCCCCGCCGGGTCACCCGG
H3_4F(XhoI)	GGCACTCGAGCGTACGCTGC
PYK1plox71F	GATCCGCGGCCGCTACCGTTCGTATAGCATACATTATACGA AGTTATCTTCCAAGTGATTTCCTTTCCTTCCC
PYK1pNcoI R	GGGCCCATGGTGTGATGATGTTTTATTTGTTTTGATTGGTG

Table A2 Plasmids in Chapter 2

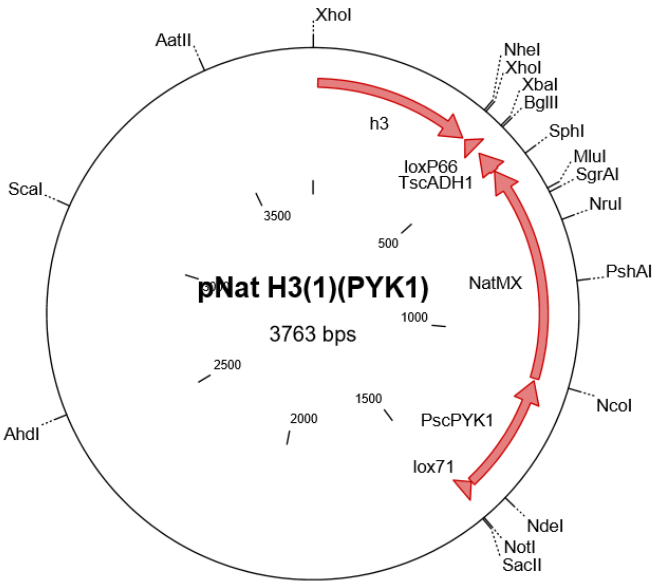
Plasmids	Construction
pRS416-SET1p-3FLAG-Set1	PCR of 3FLAG-SET1 using Set1 260F and Set1-262R, the PCR product was digested by SpeI and HindIII, and ligated to SpeI and HindIII sites of pRS416. The plasmid was mutagenesis PCR by Set1 276 and 277 to loop out loxP site.
pRS416-SET1p-3FLAG-Set1 (1-375)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 298 and 299
pRS416-SET1p-3FLAG-Set1 (379-1080)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 278 and 279
pRS416-SET1p-3FLAG-Set1 (1-689)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 294 and 295
pRS416-SET1p-3FLAG-Set1 (690-1080)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 282 and 283
pRS416-SET1p-3FLAG-Set1 (51-1080)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 284 and 285
pRS416-SET1p-3FLAG-Set1 (102-1080)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 286 and 287
pRS416-SET1p-3FLAG-Set1 (151-1080)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 288 and 289
pRS416-SET1p-3FLAG-Set1 (211-1080)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 290 and 291
pRS416-SET1p-3FLAG-Set1 Δ (101-151)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 318 and 319
pRS416-SET1p-3FLAG-Set1 (RP->AA)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 320 and 321
pRS416-SET1p-3FLAG-Set1 (R133E)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 330 and 331
pRS416-SET1p-3FLAG-Set1 (S136A)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 332 and 333
pRS416-SET1p-3FLAG-Set1 (KFH->AAA)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 324 and 325
pRS416-SET1p-3FLAG-Set1 (YFD->AAA)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 326 and 327
pRS416-SET1p-3FLAG-Set1 Δ (KFHYFD)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 322 and 323
pRS416-SET1p-3FLAG-Set1 (H150A)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 338 and 339
pRS416-SET1p-3FLAG-Set1 (Y151A)	Mutagenesis of pRS416-SET1p-3FLAG-Set1 using Set1 336 and 337

Table A2 continued

pRS416-PYK1p-3FLAG-Set1	PCR of 3FLAG-SET1 using Set1 261F and Set1-262R, the PCR product was digested by BglII and HindIII, and ligated to BamHI and HindIII sites of pRS416-PYK1.
pRS416-PYK1p-3FLAG-Set1 Δ (230-335)	Mutagenesis of pRS416-PYK1p-3FLAG-Set1 using Set1 300 and 301
pRS425-PYK1p-3FLAG-Set1 (1-375)	PCR of 3FLAG-SET1 using Set1 261F and Set1-263R, the PCR product was digested by BglII and XhoI and ligated to BamHI and XhoI sites of pRS425-PYK1.
pRS425-PYK1p-3FLAG-Set1 (151-375)	PCR of 3FLAG-SET1 using Set1 265F and Set1-263R, the PCR product was digested by BglII and XhoI and ligated to BamHI and XhoI sites of pRS425-PYK1.
pRS425-PYK1p-3FLAG-Set1 (1-275)	PCR of 3FLAG-SET1 using Set1 261F and Set1-265R, the PCR product was digested by BglII and XhoI and ligated to BamHI and XhoI sites of pRS425-PYK1.
pRS425-PYK1p-3FLAG-Set1 (1-250)	PCR of 3FLAG-SET1 using Set1 261F and Set1-264R, the PCR product was digested by BglII and XhoI and ligated to BamHI and XhoI sites of pRS425-PYK1.
pRS415-PYK1p-NLS-GST	PCR of pGEX using NLS-GST F and GST-TEV R, the PCR product digested by XbaI and BamHI, and ligated in to pRS415-PYK1 to generate pRS41-PYK1p-NLS-GST
pRS415-PYK1p-NLS-GST-CTD	PCR from BY4741 genomic DNA using CTD F and CTD 1R, the PRC product digested by BamHI and XhoI, and ligated in to the BamHI and XhoI sites of pRS415-NLS-GST
pRS415-PYK1p-NLS-GST-CTD S5A	PCR from synthesized S5A plasmid using S5A F and CTD 2R, the PRC product digested by BamHI and XhoI, and ligated in to the BamHI and XhoI sites of pRS415-NLS-GST
pRS415-PYK1p-NLS-GST-CTD S2A	PCR from synthesized S5A plasmid using S2A F and CTD 2R, the PRC product digested by BamHI and XhoI, and ligated in to the BamHI and XhoI sites of pRS415-NLS-GST

Table A3 Plasmids for generating histone mutants

pKan-H3(1)	PCR of synthesized H3(1) gBlock using H3_3F and H3 R, digested by XhoI and BglII; PCR of pK3F using PTEF F and TTEF R, digested by NotI and BglII; and the two fragments ligated to the XhoI and NotI sites of pK3FS.
pNat-H3(1)	PCR of pN3FS-CYC1 using PTEF F and NatMX-001R, digested by NotI and BglII, and ligated to BglII, and NotI sites of pKan-H3(1)
pHyg-H3(1)	PCR of pH3FS using PTEF F and HphMX-001R, digested by NotI and BglII, and ligated to BglII, and NotI sites of pKan-H3(1)
pHyg-H3(2)	PCR of synthesized H3(2) gBlock using H3_4F and H3 R, digested by XhoI and BglII, and ligated to the XhoI and BglII sites of pHyg-H3(1)
pNat-H3(1) (PYK1p)	PCR of PYK1lox71 F and PYK1 NcoI R, the PCR product digested by NotI and NcoI, and ligated to the NotI and NcoI sites of pNAT-H3(1)
pKan-lox66-71	XhoI digestion of pKan-H3(1) and self-ligation.
pHyg-lox66-71	XhoI digestion of pHyg-H3(1) and self-ligation.
pNAT-lox66-71	XhoI digestion of pNAT-H3(1) and self-ligation.



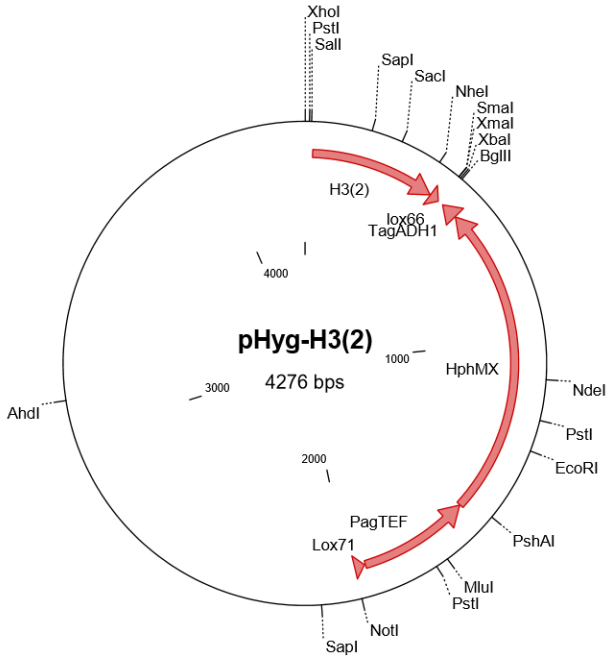

```
    XhoI
1   ctcgagggaa gcatggctag gaccaaacag actgcgagga aatctaccgg cggaaaggct cccaggaaac agtttggcctc caaggccgct aggaagtctg caccctccac
    gagctccctt cgtaccgatc ctggtttgtc tgacgctcct ttagatggcc gcctttccga gggtcctttg tcaaccggag gttccggcga tccttcagac gtgggaggtg
                >>.....................................................h3.....................................................>
                  m  a  r  t  k  q  t  a  r  k  s  t  g  g  k  a  p  r  k  q  l  a  s  k  a  a  r  k  s  a  p  s

111 tggcggggtc aaaaagccgc atcgttacaa acctggcacc gtcgccctaa gggagattag gcgttttcag aagtccaccg agctactaat aaggaaacta ccctttcaga
    accgccccag ttttcggcg tagcaatgtt tggaccgtgg cagcgggatt ccctctaatc cgcaaaagtc ttcaggtggc tcgatgatta ttcctttgat gggaaagtct
    >.....................................................h3.....................................................>
    t  g  g  v  k  k  p  h  r  y  k  p  g  t  v  a  l  r  e  i  r  r  f  q  k  s  t  e  l  l  i  r  k  l  p  f  q

221 ggcttgttag ggagattgcg caggacttta aaactgatct acgtttccag tcctccgcta ttggcgctct tcaggagtcc gttgaggcat atttggtttc cttgttcgag
    ccgaacaatc cctctaacgc gtcctgaaat tttgactaga tgcaaaggtc aggaggcgat aaccgcgaga agtcctcagg caactccgta taaaccaaag gaacaagctc
    >.....................................................h3.....................................................>
     r  l  v  r  e  i  a  q  d  f  k  t  d  l  r  f  q  s  s  a  i  g  a  l  q  e  s  v  e  a  y  l  v  s  l  f  e

                                                                                                    XhoI
                                                                                              NheI
331 gacactaatc tagctgcaat ccatgctaaa agggttacca tacagaagaa agacataaaa ttagccaggc gtctgcgtgg cgagaggagc tagctcgagt accgttcgta
    ctgtgattag atcgacgtta ggtacgattt tcccaatggt atgtcttctt tctgtatttt aatcggtccg cagacgcacc gctctcctcg atcgagctca tggcaagcat
    >.....................................................h3.....................................................>>
      d  t  n  l  a  a  i  h  a  k  r  v  t  i  q  k  k  d  i  k  l  a  r  r  l  r  g  e  r  s  -
                                                                                                          <<.loxP66..<
                                                                                                             g  n  t
```

Figure A3 Plasmid map of pNat-H3(1)(PYK1) and H3(1) sequence

```
LOCUS       pNat H3(1)   3763 bp    DNA   CIRCULAR SYN       06-DEC-2016
DEFINITION  Ligation of inverted Fragment 2 into Fragment 2
ACCESSION   pNat H3(1)
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 3763)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by SciEd Central, Scientific & Educational Software
COMMENT     SECNOTES|Vector molecule:  Fragment 2
              Fragment ends: NotI and NcoI
              Fragment size: 3410
            Insert molecule:  Fragment 2
              Fragment ends: NotI and NcoI
              Fragment size: 353
FEATURES             Location/Qualifiers
     CDS             13..423
                     /gene="h3"
                     /SECDrawAs="Gene"
     CDS             complement (430..463)
                     /gene="loxP66"
                     /SECDrawAs="Gene"
     CDS             complement (481..541)
                     /gene="TscADH1"
                     /SECDrawAs="Gene"
     CDS             complement (542..1111)
                     /gene="NatMX"
                     /product="Nourseothricin acetyltransferase"
                     /SECDrawAs="Gene"
     CDS             complement (1117..1428)
                     /gene="PscPYK1"
                     /SECDrawAs="Gene"
     CDS             complement (1429..1462)
                     /gene="lox71"
                     /SECDrawAs="Gene"
BASE COUNT      951 a    938 c    981 g    893 t
ORIGIN
        1 ctcgagggaa gcatggctag gaccaaacag actgcgagga aatctaccgg cggaaaggct
       61 cccaggaaac agttggcctc caaggccgct aggaagtctg caccctccac tggcggggtc
      121 aaaaagccgc atcgttacaa acctggcacc gtcgccctaa gggagattag gcgttttcag
      181 aagtccaccg agctactaat aaggaaacta ccctttcaga ggcttgttag ggagattgcg
      241 caggacttta aaactgatct acgtttccag tcctccgcta ttggcgctct tcaggagtcc
      301 gttgaggcat atttggtttc cttgttcgag gacactaatc tagctgcaat ccatgctaaa
      361 agggttacca tacagaagaa agacataaaa ttagccaggc gtctgcgtgg cgagaggagc
```

421 tagctcgagt accgttcgta taatgtatgc tatacgaagt tatcccgggt ctagagatct
481 atacacttat ttttttttata acttatttaa taataaaaat cataaatcat aagaaattcg
541 cttaggggca gggcatgctc atgtagagcg cctgctcgcc gtccgaggcg gtgccgtcgt
601 acagggcggt gtccaggccg cagagggtga accccatccg ccggtacgcg tggatcgccg
661 gtgcgttgac gttggtgacc tccagccaga ggtgcccggc gccccgctcg cgggcgaact
721 ccgtcgcgag ccccatcaac gcgcgcccga ccccgtgccc ccggtgctcc ggggcgacct
781 cgatgtcctc gacggtcagc cggcggttcc agccggagta cgagacgacc acgaagcccg
841 ccaggtcgcc gtcgtccccg tacgcgacga acgtccggga gtccgggtcg ccgtcctccc
901 cgtcgtccga ttcgtcgtcc gattcgtcgt cggggaacac cttggtcagg ggcgggtcca
961 ccggcacctc ccgcagggtg aagccgtccc cggtggcggt gacgcggaag acggtgtcgg
1021 tggtgaagga cccatccagt gcctcgatgg cctcggcgtc ccccgggaca ctggtgcggt
1081 accggtaagc cgtgtcgtca agagtggtac ccatggtgtg atgatgtttt atttgttttg
1141 attggtgtct tgtaaataga aacaagagag aataataaac aagttaagaa taaaaaacca
1201 aaggatgaaa aagaatgaat atgaaaaaga gtagagaata actttgaaag gggaccatga
1261 tataactgga aaaaagaggt tcttggaaat gaaaagttac caaagagtat ttataattca
1321 gaaaaaaaag ccaacgaata tcgttttgat ggcgagcctt ttttttttt taggaagaca
1381 ctaaaggtac ctagcatcat atgggaagga aaggaaatca cttggaagat aacttcgtat
1441 aatgtatgct atacgaacgg tagcggccgc ggatctgccg gtctccctat agtgagtcgt
1501 attaatttcg ataagccagg ttaacctgca ttaatgaatc ggccaacgcg cggggagagg
1561 cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc gctcggtcgt
1621 tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat ccacagaatc
1681 aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa
1741 aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa
1801 tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc
1861 ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc
1921 cgcctttctc ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta ggtatctcag
1981 ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaacccccgg ttcagcccga
2041 ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc
2101 gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac
2161 agagttcttg aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg
2221 cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca
2281 aaccaccgct ggtagcggtg gttttttttgt ttgcaagcag cagattacgc gcagaaaaaa
2341 aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa
2401 ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt
2461 aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag
2521 ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat
2581 agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac catctggccc
2641 cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa
2701 ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg cctccatcca
2761 gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa
2821 cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta tggcttcatt
2881 cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc
2941 ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact
3001 catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc
3061 tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg
3121 ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgct

3181 catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc
3241 cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta ctttcaccag
3301 cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa taagggcgac
3361 acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca tttatcaggg
3421 ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt
3481 tccgcgcaca tttccccgaa aagtgccacc tgacgtctaa gaaaccatta ttatcatgac
3541 attaacctat aaaaataggc gtatcacgag gccctttcgt ctcgcgcgtt tcggtgatga
3601 cggtgaaaac ctctgacaca tgcagctccc ggagacggtc acagcttgtc tgtaagcgga
3661 tgccgggagc agacaagccc gtcagggcgc gtcagcgggt gttggcgggt gtcggggctg
3721 gcttaactat gcggcatcag agcagattgt actgagagtg cac
//




```
                  XhoI          PstI                SalI
  1  ctcgagcgta cgctgcaggt cgacatggca cgtacgaagc agaccgcccg taagagtacg ggaggcaaag cacctcgtaa gcagcttgca agtaaagcag cgcgtaagag
     gagctcgcat gcgacgtcca gctgtaccgt gcatgcttcg tctggcgggc attctcatgc cctccgtttc gtggagcatt cgtcgaacgt tcatttcgtc gcgcattctc
                              >>.....................................H3(2).....................................>
                                m  a  r  t  k  q  t  a  r  k  s  t  g  g  k  a  p  r  k  q  l  a  s  k  a  a  r  k

                                                                                                  SapI
111  cgctcctagc acaggaggcg taaaaaaacc acataggtac aaacccggaa cagtagcatt acgtgagata cgtaggtttc agaagagcac agagcttctt attcgtaaac
     gcgaggatcg tgtcctccgc atttttttgg tgtatccatg tttgggcctt gtcatcgtaa tgcactctat gcatccaaag tcttctcgtg tctcgaagaa taagcatttg
     >.....................................................H3(2).....................................................>
     s  a  p  s  t  g  g  v  k  k  p  h  r  y  k  p  g  t  v  a  l  r  e  i  r  r  f  q  k  s  t  e  l  l  i  r  k

                                                                SacI
221  ttccatttca gcgtctagta cgtgagattg cccaggactt taaaacagat ttaaggttcc agagctcagc aataggagca ttacaggaat ctgtagaggc gtatctagta
     aaggtaaagt cgcagatcat gcactctaac gggtcctgaa attttgtcta aattccaagg tctcgagtcg ttatcctcgt aatgtccttа gacatctccg catagatcat
     >.....................................................H3(2).....................................................>
      l  p  f  q  r  l  v  r  e  i  a  q  d  f  k  t  d  l  r  f  q  s  s  a  i  g  a  l  q  e  s  v  e  a  y  l  v

                                                                          NheI
331  agtctattcg aggatacaaa cttagcagcg atacatgcaa aaagagtaac aattcagaaa aaagacataa agctagcacg taggttgagg ggagagcgtt cttgataccg
     tcagataagc tcctatgttt gaatcgtcgc tatgtacgtt tttctcattg ttaagtcttt tttctgtatt tcgatcgtgc atccaactcc cctctcgcaa gaactatggc
     >.....................................................H3(2).....................................................>>
       s  l  f  e  d  t  n  l  a  a  i  h  a  k  r  v  t  i  q  k  k  d  i  k  l  a  r  r  l  r  g  e  r  s  -
                                                                                                          lox66 >>..>
                                                                                                                y
```

Figure A4 Plasmid map of pHyg-H3(2) and H3(2) sequence

```
LOCUS       pHyg-H3(2)   4276 bp    DNA    CIRCULAR SYN       06-DEC-2016
DEFINITION  Ligation of Fragment 2 into Fragment 3
ACCESSION   pHyg-H3(2)
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 4276)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by SciEd Central, Scientific & Educational Software
COMMENT     SECNOTES|Vector molecule:  Fragment 3
              Fragment ends: BglII and XhoI
              Fragment size: 3796
            Insert molecule:  Fragment 2
              Fragment ends: XhoI and BglII
              Fragment size: 480
FEATURES             Location/Qualifiers
     CDS             25..435
                     /gene="H3(2)"
                     /SECDrawAs="Gene"
     CDS             436..469
                     /gene="lox66"
                     /SECDrawAs="Gene"
     CDS             complement (487..543)
                     /gene="TagADH1"
                     /SECDrawAs="Gene"
     CDS             complement (544..1572)
                     /gene="HphMX "
                     /SECDrawAs="Gene"
     CDS             complement (1573..1939)
                     /gene="PagTEF"
                     /SECDrawAs="Gene"
     CDS             complement (1942..1975)
                     /gene="Lox71"
                     /SECDrawAs="Gene"
BASE COUNT     1073 a   1090 c   1109 g   1004 t
ORIGIN
        1 ctcgagcgta cgctgcaggt cgacatggca cgtacgaagc agaccgcccg taagagtacg
       61 ggaggcaaag cacctcgtaa gcagcttgca agtaaagcag cgcgtaagag cgctcctagc
      121 acaggaggcg taaaaaaacc acataggtac aaacccggaa cagtagcatt acgtgagata
      181 cgtaggtttc agaagagcac agagcttctt attcgtaaac ttccatttca gcgtctagta
      241 cgtgagattg cccaggactt taaaacagat ttaaggttcc agagctcagc aataggagca
      301 ttacaggaat ctgtagaggc gtatctagta agtctattcg aggatacaaa cttagcagcg
```

361 atacatgcaa aaagagtaac aattcagaaa aaagacataa agctagcacg taggttgagg
421 ggagagcgtt cttgataccg ttcgtataat gtatgctata cgaagttatc ccgggtctag
481 agatctaaaa gtcaggcgat cacacattaa aagctataca ttacaaagcg ttaatagcgg
541 gcattattcc tttgccctcg gacgagtgct ggggcgtcgg tttccactat cggcgagtac
601 ttctacacag ccatcggtcc agacggccgc gcttctgcgg gcgatttgtg tacgcccgac
661 agtcccggct ccggatcgga cgattgcgtc gcatcgaccc tgcgcccaag ctgcatcatc
721 gaaattgccg tcaaccaagc tctgatagag ttggtcaaga ccaatgcgga gcatatacgc
781 ccggagccgc ggcgatcctg caagctccgg atgcctccgc tcgaagtagc gcgtctgctg
841 ctccatacaa gccaaccacg gcctccagaa gaagatgttg gcgacctcgt attgggaatc
901 cccgaacatc gcctcgctcc agtcaatgac cgctgttatg cggccattgt ccgtcaggac
961 attgttggag ccgaaatccg cgtgcacgag gtgccggact tcggggcagt cctcggccca
1021 aagcatcagc tcatcgagag cctgcgcgac ggacgcactg acggtgtcgt ccatcacagt
1081 ttgccagtga tacacatggg gatcagcaat cgcgcatatg aaatcacgcc atgtagtgta
1141 ttgaccgatt ccttgcggtc cgaatgggcc gaacccgctc gtctggctaa gatcggccgc
1201 agcgatcgca tccatggcct ccgcgaccgg ctgcagaaca gcgggcagtt cggtttcagg
1261 caggtcttgc aacgtgacac cctgtgcacg gcgggagatg caataggtca ggctctcgct
1321 gaattcccca atgtcaagca cttccggaat cgggagcgcg gccgatgcaa agtgccgata
1381 aacataacga tctttgtaga aaccatcggc gcagctattt acccgcagga catatccacg
1441 ccctcctaca tcgaagctga aagcacgaga ttcttcgccc tccgagagct gcatcaggtc
1501 ggagacgctg tcgaactttt cgatcagaaa cttctcgaca gacgtcgcgg tgagttcagg
1561 ctttttaccc atggttgttt atgttcggat gtgatgtgag aactgtatcc tagcaagatt
1621 ttaaaaggaa gtatatgaaa gaagaacctc agtggcaaat cctaaccttt tatatttctc
1681 tacaggggcg cggcgtgggg acaattcaac gcgtctgtga ggggagcgtt tccctgctcg
1741 caggtctgca gcgaggagcc gtaatttttg cttcgcgccg tgcggccatc aaaatgtatg
1801 gatgcaaatg attatacatg gggatgtatg ggctaaatgt acgggcgaca gtcacatcat
1861 gcccctgagc tgcgcacgtc aagactgtca aggagggtat tctgggcctc catgtcgctg
1921 gccgggtgac ccggcgggga cataacttcg tataatgtat gctatacgaa cggtagcggc
1981 cgcggatctg ccggtctccc tatagtgagt cgtattaatt tcgataagcc aggttaacct
2041 gcattaatga atcggccaac gcgcggggag aggcggtttg cgtattgggc gctcttccgc
2101 ttcctcgctc actgactcgc tgcgctcggt cgttcggctg cggcgagcgg tatcagctca
2161 ctcaaaggcg gtaatacggt tatccacaga atcaggggat aacgcaggaa agaacatgtg
2221 agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca
2281 taggctccgc ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa
2341 cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc
2401 tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc
2461 gctttctcaa tgctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct
2521 gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg
2581 tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag
2641 gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt ggcctaacta
2701 cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg
2761 aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt
2821 tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt
2881 ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt tggtcatgag
2941 attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat
3001 ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc

```
3061 tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgtagat
3121 aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac cgcgagaccc
3181 acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag
3241 aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag
3301 agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta caggcatcgt
3361 ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg
3421 agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt
3481 tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac tgcataattc
3541 tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc
3601 attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa
3661 taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg
3721 aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca ctcgtgcacc
3781 caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag
3841 gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt
3901 cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt
3961 tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc
4021 acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac
4081 gaggcccttt cgtctcgcgc gtttcggtga tgacggtgaa aacctctgac acatgcagct
4141 cccggagacg gtcacagctt gtctgtaagc ggatgccggg agcagacaag cccgtcaggg
4201 cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac tatgcggcat cagagcagat
4261 tgtactgaga gtgcac
//
```

VITA

VITA

YUEPING ZHANG

Department of Biochemistry
Purdue University
175 South University St, West Lafayette IN, 47907

Education

2009-present GPA: 3.82	PhD candidate in Biochemistry, Department of Biochemistry, Purdue University, West Lafayette, IN, USA Advisor: Dr. Scott Briggs
2005-2009	B.S. in Biological Science, College of Life science, Peking University, Beijing, China. Advisor: Dr. Xiaofeng Zheng

Publication

Zhang YP., Serratore ND., Briggs SD., N-ICE plasmids for generating N-terminal 3×FLAG tagged genes that allow Inducible, Constitutive, or Endogenous expression in *Saccharomyces cerevisiae*, 2016, Yeast (Accepted)

Zhang YP., Serratore ND., Briggs SD., Identification and Characterization of the Set1 RNA Polymerase II Interacting Motif (SRIM) in Set1 Recruitment and Histone H3K4 Methylation Deposition (In Preparation)

Honors

2015	Bird Stair Fellowship
2007	China Undergraduate Innovation Experiment Project
2004	China Physics Olympiad (ChPO), First Prize, 11st in Liaoning Province, China

Teaching Experience

Teaching assistant and instructor, BCHM309 Biochemistry Lab, Dr. Orla Hart, Spring and Summer 2016,

Teaching assistant and instructor, BIOL232 Laboratory in Biology III: cell structure and function, Dr. John Anderson, Fall 2010
Teaching assistant, BCHM462 Metabolism, Dr. Scott Briggs, Fall 2011, Fall 2012, Fall 2013
Teaching assistant, BCHM562 Biochemistry II, Dr. Xiaoqi Liu, Spring 2012, Spring 2013
Teaching assistant, BCHM695 Gene Expression, Dr. Elizabeth Tran, Spring 2014
Teaching assistant, BCHM695 Pathways, Dr. Xiaoqi Liu, Fall 2014
Teaching assistant, BCHM610 Gene Expression, Dr. Scott Briggs, Spring 2015

Poster and Oral Presentations:

2011-2016	Biochemistry Annual Retreat, Purdue University, West Lafayette, Indiana
2014	2014 midwest yeast meeting, Northwestern University, Evanston, Illinois
2014-2015	Purdue Center for Cancer Research Annual Scientific Retreat
2011-2016	BCHM Grad-Postdoc Seminars Presentations