

**REGULATORY MECHANISMS FOR THE POL II ASSOCIATED HISTONE  
METHYLTRANSFERASE SET2**

APPROVED BY SUPERVISORY COMMITTEE

Bing Li, PhD

---

Michael Roth, PhD

---

William Kraus, PhD

---

Yi Liu, PhD

---

## **DEDICATION**

To my parents, for their love and support

To my mentor, committee members and colleagues for their guidance

**REGULATORY MECHANISMS FOR THE POL II ASSOCIATED HISTONE  
METHYLTRANSFERASE SET2**

By

YI WANG

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2014

# **REGULATORY MECHANISMS FOR THE POL II ASSOCIATED HISTONE METHYLTRANSFERASE SET2**

Yi Wang

The University of Texas Southwestern Medical Center at Dallas, 2014

Supervising Professor: Bing Li Ph.D.

## **Summary**

Nucleosomes are building blocks of the eukaryotic chromatin which package genomic DNA with histones. The modification patterns of histones constitute an important signaling pathway for various nuclear processes. H3K36 methylation is catalyzed by the histone methyltransferase Set2 during transcription elongation. This important histone mark is ubiquitously presented in all organisms from yeast to mammals. In this study, we set out to investigate the molecular mechanisms by which the Set2 activity is precisely regulated during dynamic transcription cycle.

In the first part of the study we discovered a novel role of the Set2 SRI domain, which is responsible for the binding of Set2 to elongating RNA polymerase II. We show that SRI also binds to DNA which determines the substrate specificity of Set2. In addition, we identified a novel auto-inhibitory role for the middle region of Set2 in regulating catalytic activity of Set2. Remarkably, mutations at this region cause hyperactivities, which in turn lead to synthetic phenotype with an essential histone chaperone FACT. Our data suggests that a temporal control for dynamic chromatin regulation is needed during transcription elongation process.

In the second part, we investigated the molecular mechanism by which elongating Pol II regulated the Set2 activity beyond its initial recruitment. Surprisingly, we found the excessive amount of phosphorylated serine residues on Pol II CTD inhibited Set2 activity in vitro. Subsequent biophysical examination revealed that the additional phosphorylated CTD repeats collaterally occupied the surface of SRI where SRI contacts DNA. Finally, we determined that the minimal recognition unit of Set2 on fully phosphorylated CTD tail is three heptad repeats,

and demonstrated that this minimal unit is sufficient for the Set2 recruitment without disrupting its catalytic activity. Since Pol II CTD utilizes repeating sequence as a scaffold for multiple factors, our results implicate that an organized spatial arrangement of these factors along CTD is necessary for accommodating their individual functions.

In the last part of this work, we examined the state-specific functions of H3K36 methylation. By manipulating the catalytic domain of Set2, we obtained two mutants that can catalyze specific methyl-states of H3K36 both in vitro and in vivo. Genetic studies showed that cells carrying these two mutants displayed distinct phenotypes in several functional pathways, including histone chaperone, CTD proline isomerization and double-strand DNA repair. Our results suggest individual methyl-state of H3K36 plays non-redundant biological roles in cells.

In summary, we discovered multiple mechanisms by which Set2 is dynamically regulated by elongating Pol II. Setd2, the human homolog of yeast Set2, has been shown recently to be one of the most important tumor suppressors among chromatin regulators. Given the highly conserved nature of this histone methyltransferase family, we believe that our mechanistic studies here may shed lights on the roles of Setd2 in tumorigenesis.

## Table of Contents

|   |     |
|---|-----|
| DEDICATION .....  | iii |
| Summary .....   | v   |
| Table of Contents .....   | vii |
| List of Figures .....   | x   |
| List of Tables .....  | xii |
| Chapter1 INTRODUCTION.....  | 2   |
| 1.1 Chromatin regulation .....  | 2   |
| 1.2 Regulation of Set2 -mediated H3K36methylation.....  | 4   |
| 1.2.1 H3K36 methyltransferases in yeast and higher eukaryotes .....   | 4   |
| 1.2.2 Factors that influence Set2 mediated HK36 methylation .....   | 5   |
| 1.3 Functions of Set2 mediated H3K36me .....  | 7   |
| 1.3.1. Molecular functions of H3K36 methylation .....   | 7   |
| 1.3.2 Implications in development and diseases .....  | 10  |
| 1.4 Pol II CTD code .....   | 11  |
| 1.4.1 Structure of the Pol II CTD.....  | 11  |
| 1.4.2 Amino acid sequence code.....   | 11  |
| 1.4.3 Post-translational modifications of CTD .....   | 13  |
| 1.4.4 CTD phosphorylation patterns within transcription cycle.....  | 14  |
| 1.4.5 CTD interaction domain (CID).....   | 15  |
| 1.5 Summary .....   | 16  |
| Chapter2 AN AUTOINHIBITORY MECHANISM DIRECTS THE HISTONE<br>METHYLTRANSFERASE SET2 TO METHYLATE ONLY TRANSCRIBED<br>CHROMATIN.....    | 18  |
| Abstract .....  | 18  |
| Introduction.....   | 19  |
| Results.....  | 22  |
| Engaging to phosphorylated CTD plays additional roles in Set2-mediated H3K36me3<br>beside stabilization and recruitment of Set2. .... | 22  |
| Full-length Set2 and Set2 SET domain display distinct substrate specificities. ....   | 23  |
| Set2 contains an auto-inhibitory domain for the H3K36 methyltransferase activity.....   | 24  |
| The middle region of Set2 intrinsically suppresses the activity of the SET domain. ....   | 25  |
| AID interferes with the binding of SET to histone H3. ....  | 26  |
| AID plays important regulatory roles in Set2 activity <i>in vitro</i> and <i>in vivo</i> . ....                                       | 27  |
| Mis-regulation of H3K36me leads to synthetic phenotypes between AID mutants and<br>histone chaperone mutant. ....                     | 29  |
| Discussion .....  | 31  |

|  |            |
|--|------------|
| Materials and Methods.....   | 47         |
| <b>Chapter 3 PRECISE CONTACT OF SET2 WITH POL II CTD REPEATS IS ESSENTIAL FOR ITS HISTONE METHYLTRANSFERASE ACTIVITY .....</b> | <b>59</b>  |
| Abstract .....   | 59         |
| Introduction.....  | 60         |
| Results.....   | 64         |
| Set2 middle domain regulates the binding of SRI to phosphorylated CTD .....  | 64         |
| Phosphorylated RNA Pol II complex inhibits Set2 activity.....  | 64         |
| Ser2 Phosphorylated CTD peptides do not inhibit Set2 activity .....  | 65         |
| Phosphorylated GST-CTD inhibits Set2 activity through the SRI domain. ....   | 66         |
| Binding to phosphorylated GST-CTD neutralizes all SRI functions.....   | 67         |
| Extra copies of phosphorylated heptapeptide CTD collaterally occupies at the DNA binding surface of SRI. ....                  | 68         |
| Extra copies of phosphorylated heptapeptide CTD inhibits Set2 HMT activity. ....   | 69         |
| Discussion .....   | 70         |
| Materials and Methods.....   | 86         |
| <b>Chapter 4 THE SPECIFIC STATES OF H3K36 METHYLATION REGULATES DIFFERENT CHROMATIN FUNCTIONS .....</b>                        | <b>99</b>  |
| Abstract .....   | 99         |
| Introduction.....  | 100        |
| Results.....   | 104        |
| Mutations at the catalytic sites of Set2 alter the specificity of Set2-mediated H3K36 methylation <i>in vivo</i> .....         | 104        |
| Set2 catalytic pocket mutants show intrinsic state-specificities of H3K36 methylation <i>in vitro</i> . ....                   | 106        |
| Set2 Y149F (me3 only) and F234Y (me3 deficient) mutants do not alter their p-Pol II binding properties. ....                   | 107        |
| Set2 catalytic pocket mutants show cryptic transcription phenotype. ....   | 107        |
| Set2 state-specific mutants display different synthetic phenotypes with histone chaperone yFACT mutant. ....                   | 108        |
| Set2 state-specific mutants display different synthetic phenotypes with prolyl-isomerase Ess1 mutant.....                      | 109        |
| DNA Double-strand break repair requires specific states of H3K36me.....  | 110        |
| Discussion .....   | 111        |
| Materials and Methods.....   | 121        |
| <b>Chapter 5 SUMMARY AND FUTURE DIRECTIONS.....</b>  | <b>129</b> |
| 5.1 Novel insights into H3K36 regulation .....   | 129        |
| 5.2 Explore the orchestration of PCAPs on p-CTD.....   | 130        |
| 5.3 Looking for H3K36 methyl-state-specific readers. ....  | 131        |

Bibliography.....133

## List of Figures

|   |    |
|---|----|
| Figure 2.1 Engaging to phosphorylated CTD plays additional roles in Set2-mediated H3K36me3 beside stabilization and recruitment of Set2. .... | 33 |
| Figure 2.2 Binding to phosphorylated Pol II CTD is essential for Set2-mediated H3K36me3. ....   | 35 |
| Figure 2.3 Full-length Set2 and Set2 SET domain display distinct substrate specificities.....   | 36 |
| Figure 2.4 Set2 contains an auto-inhibitory domain for the H3K36 methyltransferase activity .....   | 38 |
| Figure 2.5 The middle region of Set2 intrinsically suppresses the activity of the SET domain.....   | 40 |
| Figure 2.6 AID interferes with the binding of SET to histone H3.....  | 41 |
| Figure 2.7 AID plays important regulatory roles in Set2 activity in vitro and in vivo. ....   | 42 |
| Figure 2.8 AID mutations render Set2 more sensitive to degradation.....   | 43 |
| Figure 2.9 AID regulates the binding of Set2 to phosphorylated CTD .....  | 44 |
| Figure 2.10 Mis-regulation of H3K36me leads to synthetic phenotypes between AID mutants and histone chaperone mutant.....                     | 45 |
| Figure 2.11 AID mutations lead to cryptic transcription phenotype.....  | 46 |
| Figure 3.1 Set2 middle domain regulates the binding of SRI to phosphorylated CTD.....   | 72 |
| Figure 3.2 Phosphorylated RNA Pol II complex inhibits Set2 activity. ....   | 73 |
| Figure 3.3 Ser2 Phosphorylated CTD peptides (4X) do not inhibit Set2 activity. ..   | 75 |
| Figure 3.4 A clone strategy to engineer mutations at any position along CTD.....  | 76 |
| Figure 3.5 Phosphorylated GST-CTD inhibits Set2 activity through the SRI domain. ....   | 78 |
| Figure 3.6 Binding to phosphorylated GST-CTD neutralizes all SRI functions. ....  | 80 |

|   |     |
|---|-----|
| Figure 3.7 Phosphorylated GST-CTD inhibits the binding of Set2 to DNA and nucleosomes. ....                                       | 81  |
| Figure 3.8 Extra copies of phosphorylated heptapeptide CTD collaterally occupies at the DNA binding surface of SRI. ....          | 82  |
| Figure 3.9 Extra copies of phosphorylated heptapeptide CTD inhibits Set2 HMT activity. ....                                       | 84  |
| Figure 4.1 Set2 contains conserved residues that are predicted to control state-specificity. ....                                 | 113 |
| Figure 4.2 Mutations at the catalytic sites of Set2 alter Set2-mediated H3K36 methylation in vivo. ....                           | 114 |
| Figure 4.3 Set2 catalytic pocket mutants show intrinsic state-specificities of H3K36 methylation in vitro. ....                   | 115 |
| Figure 4.4 Set2 Y149F (H3K36me3 only) and F234Y (H3K36me3 deficient) mutants do not alter their p-Pol II binding properties. .... | 116 |
| Figure 4.5 Set2 catalytic pocket mutants cause intragenic spurious transcription of STE11. ....                                   | 117 |
| Figure 4.6 Set2 state-specific mutants display different synthetic phenotypes with histone chaperone yFACT mutant. ....           | 118 |
| Figure 4.7 Set2 state-specific mutants display different synthetic phenotypes with RNA Pol II prolyl-isomerase Ess1 mutant. ....  | 119 |
| Figure 4.8 DNA Double-strand break repair requires specific states of H3K36me. ....   | 120 |

## List of Tables

|                             |     |
|-----------------------------|-----|
| Table 2.1 Plasmid List..... | 53  |
| Table 2.2 Strain List.....  | 55  |
| Table 2.3 Primer List.....  | 55  |
| Table 3.1 Plasmid List..... | 96  |
| Table 3.2 Strain List.....  | 98  |
| Table 4.1 Plasmid List..... | 125 |
| Table 4.2 Stains List.....  | 126 |

# Chapter1 INTRODUCTION

## 1.1 Chromatin regulation

Chromatin is organized in arrays of nucleosomes (Kornberg, 1977). Two copies of each histone proteins, H2A, H2B, H3 and H4, form an octamer that is wrapped by a 146 base pairs of DNA to assemble a nucleosome core. DNA interacts with histone octamers through fourteen independent contacts (Luger and Richmond, 1998). Nucleosomes are densely packed to achieve the 10,000–20,000-fold compaction which is necessary to fit a genome into the small volume of the nucleus (Woodcock and Ghosh, 2010). To coordinate with various nuclear processes, the compacted nucleosomes also need to be dynamically regulated in several pathways, including ATP-dependent chromatin remodeling enzymes (Hargreaves and Crabtree, 2011), covalent modification of histones (Zentner and Henikoff, 2013) and replacement of canonical core histones with specialized variants (Talbert and Henikoff, 2010).

### *Chromatin remodeling and histone dynamics*

Nucleosome remodelers use the energy produced from ATP hydrolysis to change the packaging state of chromatin by mobilizing, sliding nucleosomes along the DNA template, or disassembling histone octamers from DNA. (Clapier and Cairns, 2009). There are currently four different families of chromatin remodeling complexes: SWI/SNF family, ISWI family, CHD family, and INO80 family. All four types of remodelers utilize ATP hydrolysis to alter histone-DNA contacts and share a similar ATPase domain but are also specialized for particular purposes and biological contexts.

### *Histone modification*

Histones are subject to a vast number of post-translational modifications (PTMs), such as methylation, acetylation, ubiquitination, ADP-ribosylation and sumoylation of lysine (K) residues; methylation of arginine (R) residues; and phosphorylation of serine (S) and threonine (T) residues (Bannister and Kouzarides, 2011). Histone modifications exert their effects via two main mechanisms. The first involves the direct influences of the overall chromatin structure and the second involves the modification related binding of effector molecules. For instance, histone acetylation effectively reduces the positive charge of histones, and this mechanism has the potential to disrupt

electrostatic interactions between histones and DNA (Hong et al., 1993; Workman and Kingston, 1998). A recent study shows that H4-K16Ac which has roles in transcriptional activation and the euchromatin maintenance (Suka et al., 2002) modulates both higher order chromatin structure and functional interactions between chromatin remodelers and the chromatin fiber (Shogren-Knaak et al., 2006). Meanwhile, it has been shown that  $\beta$ -globin locus where the genes reside within a hyper-acetylated and transcriptionally competent chromatin environment displays DNase sensitivity (Kouzarides, 2007).

Numerous chromatin-associated factors have been shown to specifically interact with modified histones via many distinct domains (Yun et al., 2011). “Histone code” hypothesis proposes that the histone modifications serve to recruit other proteins to bring about distinct downstream events (Strahl and Allis, 2000). Methylation is recognized by chromo-like domains of the Royal family (chromo, tudor and MBT) and unrelated domains such as ankyrin repeats, zf-CW, PWWP and PHD domains (Yun et al., 2011). Although different binders are folded differently to fulfill other structural requirements, binding surfaces of the domains that recognize the same mark remarkably resemble each other. For example, TAF3-PHD domain and double-Tudor domain of JMJD2A both apply similar binding surfaces to recognize H3K4me3 peptide (Huang et al., 2006). Moreover, distinct from other modifications, lysine methylation has multiple methyl states (mono-, di-, and tri- methylation). In some lysines, different methyl-states are recognized by different sets of effectors to differentiate the state-specificity of methylation. For instance, Pdp1 binds to H4K20me1 for cell-cycle regulation, whereas Crb2 recognizes H4K20me2 to control a DNA damage checkpoint (Wang and Jia, 2009). However, at other sites, methyl states only control the binding strength of the same chromatin regulators. For example Rpd3S binds to H3K36me3 nucleosomes with the highest affinity and these binding decreases as the number of methyl state reduces (Li et al., 2009a).

Acetylation is recognized by bromo domains (Dhalluin et al., 1999) and the tandem PHD domain (Lange et al., 2008; Zeng et al., 2010). Many bromo domains bind to multiple acetylated histones, and the tandem PHD domain of human DPF3b also prefers acetylated H3 and H4 (Lange et al., 2008), indicating the lack of unique sequence recognition by these readers. Meanwhile, two readers have been identified to recognize histone phosphorylation. The BRCT domain of MDC1 binds to phosphorylated Serine near the C-terminus of histone H2AX (Stucki et al., 2005). Mammalian 14-3-3 $\zeta$  recognizes H3S10ph peptide using a deep scaffold (Macdonald et al., 2005) while yeast 14-3-3 proteins Bmh1 and Bmh2 also bind to H3S10ph (Walter et al., 2008).

### ***Histone Variant Incorporation***

Histone variants have distinct amino acid sequences that can influence both the physical properties and dynamics of the nucleosome. The genome packaged by canonical histones can be replaced with histone variants that alter nucleosome structure, stability, dynamics, and, ultimately, DNA accessibility (Weber and Henikoff, 2014).

For instance, H3.3 is preferentially incorporated over transcribed regions independent of DNA replication (Mito et al., 2005). Yeast version of H3 most closely resembles the H3.3 variant rather than the replication-dependent H3.1 in human (Malik and Henikoff, 2003). In metazoans, H3.3 differs from H3 by only four to five amino acids (Filipescu et al., 2013). Three of these differences are found within the core histone fold domain and specify the alternative deposition pathways (Ahmad and Henikoff, 2002). Meanwhile, H2A.Z has been found to associate with gene activation and promote elongation (Santisteban et al., 2011). H2A.Z is only ~60% identical to H2A within species, but is relatively conserved between different species and is essential in metazoans (Zlatanova and Thakar, 2008). H2A.Z has an extended acidic patch, which stimulates remodeling activity with the ISWI ATP-dependent remodeler (Goldman et al., 2010). In yeast, H2A.Z can be bound by the general H2A/H2B chaperone Nap1 or Chz1, which preferentially binds H2A.Z over H2A (Luk et al., 2010). These chaperones provide a source of H2A.Z for the Swr1 remodeling complex, which exchanges H2A.Z for H2A (Mizuguchi et al., 2004).

## **1.2 Regulation of Set2 -mediated H3K36methylation**

Like all other histone methylation, H3 lysine 36 residues can be modified with mono-, di- or trimethyl groups, which significantly extends the complicity of the H3K36 methylation “language” (Zhang et al., 2012c). Set2-mediated H3K36me is modulated by a variety of factors including elongating Pol II, transcription factors, histone chaperones, histone binding surfaces and demethylases (Venkatesh and Workman, 2013).

### **1.2.1 H3K36 methyltransferases in yeast and higher eukaryotes**

To date, at least eight distinct mammalian enzymes have been identified to methylate H3K36 *in vitro* and /or *in vivo*. All of them have the catalytic Su(var)3-9, E(z), and trithorax (**SET**) domain in common but they also have varying preferences to produce different methylation states on lysine 36

(Wagner and Carpenter, 2012). In both budding and fission yeasts, Set2 is the only methyltransferase performing all three methylation reactions at H3K36 (Morris et al., 2005; Strahl et al., 2002). However, human Setd2/Hypb, the ortholog of yeast Set2, is thought to be the only H3K36 tri-methylase (Edmunds et al., 2008; Yuan et al., 2009). Meanwhile, other H3K36 methyltransferase also have been discovered: Nuclear receptor SET domain-containing (NSD) family members (NSD1/2/3) have been found as H3K36 mono- and di- methyltransferases both in vivo and in vitro (Kuo et al., 2011; Li et al., 2009b; Rayasam et al., 2003; Yang et al., 2008) while NSD1 and NSD2 also methylate H4K20 in different substrate contexts (Hajdu et al., 2011; Li et al., 2009b; Pei et al., 2011). Moreover, ASH1L (An et al., 2011; Gregory et al., 2007), SMYD2 ((Abu-Farha et al., 2008; Brown et al., 2006), SETMAR (Fnu et al., 2011; Lee et al., 2005) and SETD3 (Eom et al., 2011) have also been identified as H3K36 mono- or di- methyltransferases.

### **1.2.2 Factors that influence Set2 mediated H3K36 methylation**

#### ***Elongating Pol II:***

Set2 has been shown to associate with the elongating form of Pol II through the binding between its C-terminal Set2-Rpb1 Interaction Domain (SRI) (Kizer et al., 2005; Krogan et al., 2003; Li et al., 2003a; Xiao et al., 2003b). SRI domain recognizes the Ser2/Ser5 phosphorylated C-terminal domain (CTD) (Li et al., 2005; Vojnic et al., 2006). Loss of the SRI domain abolishes di- and tri-methylation of H3K36 (Kizer et al., 2005), while over-expressing SET domain alone restores H3K36me<sub>2</sub> (Youdell et al., 2008) indicating phosphorylated CTD contact is particularly required for H3K36me<sub>3</sub> (Fuchs et al., 2012).

#### ***Transcription elongating factors:***

RNA Pol II associated factor 1 (Paf1) and CTD kinase CTDK-1, Bur1/Bur2 complex have been shown to be involved in transcription elongation and histone modification regulation (Hsin and Manley, 2012; Jaehning, 2010). Disrupting each subunit of the transcription factor Paf1 complex, Paf1, Cdc73 or Ctr9 results in diminished H3K36 tri-methylation (Chu et al., 2007; Li et al., 2009a). Consistent with the fact that Set2 only binds to phosphorylated Pol II, the kinases were shown to regulate Set2 as well. Loss of the CTDK-1 kinase subunit, Ctk1 abrogates methylation of H3K36, while deletion of the other CTD kinase gene *BUR1* or *BUR2* results in the loss of H3K36 tri-methylation (Chu et al., 2006; Youdell et al., 2008).

#### ***Histone chaperones:***

Histone chaperones are key proteins that function at multiple steps of nucleosome formation (Burgess and Zhang, 2013). Multiple lines of evidence suggest histone chaperones are involved in H3K36 methylation. In yeast, it has been shown that *spt16* mutation (subunit of FACT complex) selectively reduces H3K36me3 (Chu et al., 2006; Youdell et al., 2008). The FACT complex has been shown to facilitate the reassembly of H2A-H2B dimers that are preferentially mono-ubiquitinated on histone H2B (Pavri et al., 2006). Consistent with this fact, loss of Large cells1 (Lge1), a component of the yeast Bre1-ubiquitin ligase complex, only causes loss of H3K36me3 (Li et al., 2009a). Another histone chaperone, Asf1, which is involved in transcription elongation, has been shown to be necessary for maintaining H3K36 tri-methylation without affecting mono- or di-methylation state (Lin et al., 2010). Moreover, *Spt6* mutants display allele-specific effects of H3K36 methylation: in *spt6-104* mutant only H3K36me3 is reduced, while in *spt6-1004* mutant there is no H3K36 methylation (Chu et al., 2006). Interestingly, in human, Iws1, a Spt6 interacting protein, is shown to recruit Hypb/Setd2 to RNA Pol II elongation complex and is required for H3K36me. This result indicates that a mega-complex containing histone chaperone and histone methyltransferase is required for maintaining the H3K36 methylation state of active genes (Yoh et al., 2008).

### ***Histone surface:***

Set2 has been shown as a nucleosomal H3 selective methyltransferase (Strahl et al., 2002) suggesting except recognizing H3K36, additional surfaces on nucleosome are also required for Set2 activity. Histone H4 is shown to allosterically activate Set2 (Du et al., 2008). Further analysis leads to the identification of a trans-histone methylation pathway demonstrating that H4K44 and other histone residues (such as H2A L116/L117) are necessary for positioning Set2 active site close to H3K36 on nucleosome (Du and Briggs, 2010). Meanwhile, H3P38 is also found necessary for H3K36me3 and isomerization of this residue by Fpr4 renders the conformation of H3K36 unsuitable for Set2 to methylate H3K36 (Nelson et al., 2006; Youdell et al., 2008). Moreover, deleting 20 amino acids, or substituting glutamines for lysines in the H3 tail, greatly impaired K36 methylation suggesting Set2 activity is controlled by an intra-tail interactions of nucleosome (Psathas et al., 2009).

### ***Set2 protein stability:***

A recent study also suggests Set2 protein levels influence H3K36 methylation states in yeast (Fuchs et al., 2012). The results show that the disruption of interaction between Set2 and phosphorylated CTD by

using yeast mutants defective in CTD phosphorylation at serine 2 results in a destabilization of Set2 protein levels and H3K36 methylation. The author proposes a model that free form of Set2 which is capable to di-methylate H3K36 is degraded in a proteasome-dependent manner. However to finish H3K36me<sub>3</sub>, more retention time of Set2 is required by associating with phosphorylated CTD serine residues.

### ***Demethylases of H3K36:***

Histone methylation is dynamically regulated by both histone methyltransferases and histone lysine demethylases (KDM) (Shi and Whetstine, 2007) which enzymatically remove histone methyl groups (Klose et al., 2006). Yeast H3K36 demethylases include Jhd1 and Rph1, both belonging to the Jumonji family histone demethylases (Shi and Whetstine, 2007). Rph1 demethylates both H3K36 di- and tri-substrates (Kim and Buratowski, 2007; Klose et al., 2007) while Jhd1 targets mono- and di- H3K36 (Fang et al., 2007).

## **1.3 Functions of Set2 mediated H3K36me**

### **1.3.1. Molecular functions of H3K36 methylation**

#### ***Transcriptional regulation:***

*Set2-Rpd3S pathway and cryptic transcription:* It has been shown that histone deacetylase complex (HDAC) Rpd3S recognizes H3K36 methylated nucleosome through both the chromo domain (CHD) (Keogh et al., 2005) and the plant homeobox domain (PHD) of its subunits Eaf3 and Rco1, respectively (Li et al., 2007b; Reid et al., 2004). Deletion of CHD domain of Eaf3, PHD domain of Rco1, Set2 as well as mutation of H3K36 to alanine results in generation of hyper-acetylated chromatin over open reading frames (ORFs) (Carrozza et al., 2005b; Joshi and Struhl, 2005; Keogh et al., 2005). The mis-regulated hyper-acetylation at ORFs has been found to cause at least two functional consequences including the initiation of intragenic cryptic promoters (Carrozza et al., 2005b; Li et al., 2007b) and increased homologous recombination (Pai et al., 2014) (discussed in **DNA damage repair**). Interestingly, in human Sin3B, Pf1, Mrg15, and HDAC1 associate in a stable complex that binds H3K4me<sub>3</sub>/H3K36me<sub>3</sub>-enriched nucleosomes to restrain the transcription. It has been shown that inactivation of this complex promotes increased Pol II progression within transcribed regions and subsequent increased transcription (Jelinic et al., 2011) suggesting similar mechanism as Set2-Rpd3S

pathway is applied in higher eukaryotes to regulate Pol II transcription.

The term cryptic transcription in this thesis mainly describes the presence of transcripts initiated from intragenic promoters that are usually inaccessible for assembly of the transcription machinery (Kaplan et al., 2003). This concept is different from the “pervasive” transcription of non-coding RNAs (ncRNA), which are also often referred to as cryptic transcript (Flynn and Chang, 2012). Cryptic promoters do not have unified characters in genome. The promoter of *FLO8* contains TATA box (Kaplan et al., 2003) while others, for instance *STE11* does not (Pattenden et al., 2010). These promoters are shown to be regulated independently from their individual canonical promoter but share the same components of the transcriptional machinery (Pattenden et al., 2010). They largely localize at infrequently transcribed loci (Cheung et al., 2008; Li et al., 2007c) and do not interfere with according full length transcriptions (Cheung et al., 2008; Smolle et al., 2012).

*Other factors affect cryptic transcription:* Many other factors are also found to regulate cryptic transcription. Some of them act within the Set2-Rpd3S pathway, while others act in parallel with Set2. Spt6 and Spt16 (FACT) were first identified to repress transcriptional initiation from cryptic sites (Kaplan et al., 2003). Later a systematic screening shows various factors including histone gene regulation factors, transcription elongation factors, chromatin assembly factors and remodelers, and histone modifiers are required to repress cryptic transcription (Cheung et al., 2008). Deletion of Ctk1, the catalytic subunit (Ctk1) of CTD kinase CTDK-1 complex, or Bur2, the cyclin factor within Bur1 complex, causes cryptic transcription (Chu et al., 2007; Youdell et al., 2008). In addition, deleting subunit of the other elongation factor Paf1 complex, Ctr9, causes cryptic transcription (Li et al., 2009a; Youdell et al., 2008). Furthermore, combinatorial effects are also observed in *BUR2Δ SET2Δ*, *PAF1Δ SET2Δ*, and *CTR9Δ SET2Δ* double mutants (Chu et al., 2007) and restoration of H3K36me2 could not rescue cryptic transcription by the *CTK1* deletion, indicating Set2 independent functions of these factors in repression of cryptic initiation (Youdell et al., 2008).

Mutation of histone H4 K44, H2A L116 and L117 reduce the H3K36me2/3 and cause cryptic transcription (Du and Briggs, 2010; Du et al., 2008). H2B ubiquitylation is involved in suppressing cryptic initiation of certain genes (Chandrasekharan et al., 2009; Fleming et al., 2008). Meanwhile, histone chaperones Rtt106, Asf1 and Rtt109 are identified to regulate cryptic transcription. Rtt106 deficient strain shows cryptic transcription (Silva et al., 2012). Deletion of *ASF1* or *RTT109* displays phenotype when combined with *SET2* deletion suggesting that Asf1 and Rtt109 function in concert with

Set2 pathway to repress cryptic transcription (Venkatesh et al., 2012). Furthermore, histone remodeler mutant  $\Delta ISWI$  alone causes production of low-to-moderate amounts of cryptic transcripts while  $\Delta ISWI/\Delta CHD1$  shows synthetic phenotype. Interestingly,  $\Delta SET2/\Delta ISWI/\Delta CHD1$  triple knockout further exacerbates the cryptic phenotype, implicating *ISWI*, *CHD1* and *SET2* are involved in the same pathway (Smolle et al., 2012).

**Dosage compensation:** Transcription-coupled H3K36 methylation has been shown to correlate with the enrichment of Male-specific lethal (MSL) complex linking this histone modification with dosage compensation in flies (Bell et al., 2008). To compensate for having only one X chromosome, male flies use the MSL complex to up-regulate the expression of X-linked genes by facilitating the progression of Pol II across the bodies of those loci (Larschan et al., 2011). In *setd2* mutant males, MSL complex maintains X specificity but exhibits reduced binding to target genes. Furthermore, recombinant MSL3 protein (an ortholog of yeast Eaf3) is found to bind nucleosomes marked by H3K36me3 *in vitro* (Larschan et al., 2007) suggesting the importance of H3K36 methylation in dosage compensation.

***DNA damage repair:***

H3K36methylation is also critical for the maintenance of genomic stability and response of DNA damage repair. *In budding yeast*, phospho-CTD bound Set2 alleviates cell growth defect of Methyl methanesulfonate (MMS) treated cells (Winsor et al., 2013). Set2 has been shown to genetically interact with various DNA damage repair response factors including major players in both non-homologous end joining (NHEJ) and homologous recombination (HR) pathways. In addition, Set2 regulates transcription-dependent dynamic transition between H3K36me3 to H3K36me2 at DSBs indicating the state-specific function of H3K36 methylation in DSB repair (Jha and Strahl, 2014). *In fission yeast*, H3K36 modification is shown to control the choice of DSB repair pathway. Set2-dependent H3K36 methylation reduces chromatin accessibility, reduces resection and promotes non-homologous end joining (NHEJ). Loss of Set2 increases H3K36Ac, chromatin accessibility and resection resulting in opposite phenotype from the loss of histone acetyltransferase Gcn5 (Pai et al., 2014). *In human*, Setd2 facilitates the recruitment of C-terminal Interacting Protein (CtIP) to promote DSB resection, allowing Replication Protein A (RPA) and RAD51 to bind at DNA damage sites. (Pfister et al., 2014). H3K36me3 is also required *in vivo* to recruit the mismatch recognition protein hMutS $\alpha$  (hMSH2-hMSH6) for DNA mismatch repair (Li et al., 2013). Interestingly, other H3K36 methyltransferase is also involved in DNA

damage repair pathway. DNA repair protein Metnase (also called SETMAR) which has a SET-like histone methylase domain is found to localize at an induced DSB and directly mediated the formation of H3K36me2 near the locus. H3K36me2 improved the association of NHEJ pathway DNA repair components, including NBS1 and Ku70, at DSB locus to help damage repair (Fnu et al., 2011)

***Alternative splicing:***

H3K36me3 is found to preferentially mark exons relative to introns. In addition, alternative exons have lower H3K36me3 signal than constitutive exons indicating H3K36me3 exon marking in chromatin provides a dynamic link between transcription and splicing (Kolasinska-Zwierz et al., 2009). H3K36me3 has been shown important to recruit splicing factor PTBP1 to favored exon region to determine cell specific expression of protein isoforms. (Luco et al., 2010). Mouse PC4 and SF2 interacting protein 1 (Psp1) is shown to bind to H3K36me3 nucleosomes directly through its PWWP domain and co-localizes with splicing factors to regulate alternative splicing (Pradeepa et al., 2012).

***Histone exchange:***

H3K36methylation has been shown to block histone exchange by suppressing histone chaperones binding and recruiting chromatin remodelers. It has been shown that H3K36 methylation suppresses the interaction of H3 with histone chaperones, histone exchange over coding regions and the incorporation of new acetylated histones (Venkatesh et al., 2012). Meanwhile, Isw1b chromatin-remodeling complex is specifically recruited to ORFs by H3K36 methylation through its PWWP domain of the Ioc4 subunit. Isw1b cooperates with Chd1 to regulate chromatin structure and prevent trans-histone exchange (Smolle et al., 2012).

**1.3.2 Implications in development and diseases**

In higher eukaryotes, defects in genes that maintain the levels of H3K36methylation cause developmental defects and diseases (Wagner and Carpenter, 2012). Homozygous disruption of *SETD2/HYPB* impairs H3K36me3 and results in embryonic lethality at E10.5-E11.5 with severe vascular defects (Hu et al., 2010). This result indicates Setd2/Hypb is required for embryonic vascular remodeling. More importantly, the *SETD2* gene is found among the most mutated cancer genes across 21 tumor types (Lawrence et al., 2014). Mutated *SETD2* is linked to sporadic clear renal cell carcinoma (Brugarolas, 2014; Duns et al., 2010), leukemia (MLL-associated acute myeloid leukemia (AML))(Zhu

et al., 2014), T-cell precursor acute lymphoblastic leukemia (Zhang et al., 2012b), pediatric acute lymphoblastic leukemia (Mar et al., 2014), breast cancer (Al Sarakbi et al., 2009; Newbold and Mokbel, 2010) and hemispheric high-grade gliomas (Fontebasso et al., 2013). Moreover, a histone H3K36M mutation that can inactivate Setd2 was linked to two early-onset cancers: chondroblastoma and giant cell tumor of bone (Behjati et al., 2013). However, the detailed mechanism of *SETD2* mutation related tumorigenesis remains largely unknown.

## 1.4 Pol II CTD code

### 1.4.1 Structure of the Pol II CTD

Pol II C-terminal repeat domain (CTD) of Rpb1, the largest subunit of Pol II is a repetitive domain that extends from the catalytic core of the enzyme. The CTD comprises tandem repeats of heptapeptides with the consensus sequence of Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (YSPTSPS) as well as non-repeat linker region and the rump domain (Corden, 2013b). While the sequence of the repeat is conserved in *S. cerevisiae* and mammals, the length of the CTD differs markedly, with 26 and 52 repeats, respectively. Though in the crystal structures of yeast Pol II, the CTD and an 80-amino-acid residue linker connecting to the polymerase core are not visible due to motion or disorder (Armache et al., 2005; Cramer et al., 2001), the RNA exit region is located near the last ordered residue of Rpb1, L1450, at the beginning of CTD linker (Cramer et al., 2001; Meinhart et al., 2005) suggesting that CTD is flexibly linked to the adjacent RNA exit region of the enzyme. The largely disordered nature of the free CTD allows for many different interactions with target proteins. In fully extended  $\beta$ -strand conformation, the theoretical length of the yeast CTD and linker would be  $\sim 650$  Å and 250 Å, respectively, while the core of Pol II is about 150 Å in diameter (Cramer et al., 2001). The phosphorylation of the CTD may produce a more extended structure as the negatively charged side chains will tend to repel one another (Zhang and Corden, 1991). Interestingly, one of the recent studies suggests that flexible protein structure adjacent to essential heptad sequence is required for CTD function by showing insertion of poly-alanine residues which form stable secondary structures causes lethality. Such less flexible spacing sequences are shown to compromise interactions between CTD kinase and CTD mutants (Liu et al., 2010).

### 1.4.2 Amino acid sequence code

#### *Individual amino acid code*

Each of seven amino acid within heptad is considered the letter of CTD code. Genetic studies have helped to define the roles of each amino acids within the repeating sequence. *In S. cerevisiae*, replacement of the phosphorylation sites, Ser2 or Ser5 with either Ala or Glu causes lethality (West and Corden, 1995), which is consistent with essential roles for both the phosphorylated and unphosphorylated form of the CTD. Altering the order of Ser2 and Pro3 also is lethal indicating the need for correct spacing of the heptad Pro residues. Changing the Tyr1 to Phe is also lethal arguing for a possible role for Tyr phosphorylation. (West and Corden, 1995). Moreover, substitution of Ser7 with Glu is lethal suggesting dephosphorylation of this residue is important (Zhang et al., 2012a). Lastly, substitution of Thr4 and Ser7 by Ala supports viability (Stiller et al., 2000). *In S. pombe*, Ala substitutions at Pro3, Ser5, and Pro6 were found to be lethal, but in contrast to the *S. cerevisiae* CTD, Ser2 to Ala or Tyr1 to Phe substitutions were not lethal (Schwer et al., 2012). Meanwhile, substituting Thr for Ser2 or Ser5 yielded different results. In the case of Ser2, Thr substitution is not lethal but a Ser5 to Thr is lethal indicating that at this position the presence of an extra methyl group interferes with CTD conformation or interaction with modifying enzymes. Substituting either Val or Ala at Thr4 was not lethal indicating that phosphorylation of this residue in *S. pombe* is not essential. Finally, Gly substitution at Pro3 is lethal while substitution at Pro6 is not, although these cells grow slowly (Schwer and Shuman, 2011). *In human cells*, Ser2 or Ser5 substituted with Ala causes lethality (Hintermair et al., 2012). However, only the substitution of Ser7 with Ala is viable while substitution of each repeat with Glu or Thr/Lys in alternating repeats is not viable. This indicates that phosphorylation of Ser7 may not be essential (Chapman et al., 2007). Moreover, Thr4 substituted with either Val or Ala is not able to support growth but Thr4 substituted with Ser is viable. This result is consistent with an essential function for Thr4 perhaps requiring phosphorylation (Hintermair et al., 2012).

### ***Functional unit of CTD***

The tandem nature of the CTD suggests that the heptad repeat is the “functional unit”. However, Stiller and colleagues showed that addition of a single alanine residue in each heptad repeat was lethal (Stiller and Cook, 2004) indicating the functional unit comprises more than one repeat. The functional unit of the CTD was first found as a two-heptad pair (Stiller and Cook, 2004) by showing the insertion of an alanine residue between di-heptads maintains almost normal growth phenotype. Later studies further narrowed down the sequence of minimal functional unit to one full heptad and the next three or four residues of the following repeat in both budding and fission yeasts. *In S. pombe* substituting

positions 5–7 of the distal di-heptad (YSPTSPSYSPTAAA)<sub>7</sub> had no discernible effect on growth indicating Ser5-Pro6 sequence is not required in every repeat. Similarly in *S. cerevisiae*, YSPTSPSYSP has been found as the minimal functional unit sequence (Liu et al., 2010; Schwer et al., 2012).

### *Spacing of functional units*

Tandemly repeated CTD sequences make functional units overlapping (e.g. any Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub> sequence outside the first heptad can also be utilized as last three letters of the minimal functional unit: Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub> Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>). It is thought that this global, tandemly repeated structure provides the largest number and densest packing of binding positions along CTD (Liu et al., 2010). And this feature of CTD also raises a question that whether the spacing of each functional unit affects CTD function. In budding yeast, the spacing of consecutive CTD functional units is not critical as up to five Ala residues between di-heptads is viable. However insertion of seven Ala residues between di-heptads is lethal but between tri-heptads is not (Stiller and Cook, 2004). Similar result has also been observed in fission yeast (Schwer et al., 2012). Installing blocks of two to five Ala at the distal end of each minimal function unit (e.g. YSPTSPSYSPTAA) shows deteriorated growth phenotypes which correlate with the increment of Ala number under various temperature conditions. Interestingly, substituting two Pro residues in a seven Ala restores viability of budding yeast by disrupting stable secondary structures that are formed by 7Ala sequence (Liu et al., 2010) indicating helical inserts may alter the conformation of adjacent functional unit.

### **1.4.3 Post-translational modifications of CTD**

All repeats of CTD undergo a dynamic cycle of post-translational modifications. The modification patterns are recognized by CTD interacting proteins. The interplay of the modifying enzymes together with the functional output generated by the recognizing factors gives rise to the idea of a CTD code. In this thesis we mainly focus on discussing serine phosphorylation of CTD.

The phosphorylation of Ser2 and 5 residues is the most characterized CTD modification (Corden, 2013a). Ser7 on CTD is also phosphorylated during transcription (Chapman et al., 2007). Kin28, the enzyme responsible for phosphorylating Ser5, also phosphorylates the Ser7 (Glover-Cutter et al., 2009). Instead of affecting expression of protein coding genes, Ser7A mutation is found to influence a class of Pol II transcribed genes encoding small nuclear RNAs (snRNA) (Egloff et al., 2007). In addition, phosphorylation of Tyr1 and the kinase responsible of this modification, cAbl were known long ago

(Baskaran et al., 1993). Tyr1P appears downstream of TSS, is present along coding sequences and drops until the end of gene. It has been shown that the presence of phosphorylated Tyr1 inhibits the association of termination factors to the CTD, thus preventing aberrant termination of transcription (Mayer et al., 2012). Furthermore, phosphorylated Thr4 is enriched in the body of genes while low levels are detected at polyA sites in yeast. Mutations of Thr4 do not influence yeast viability, but its substitution in higher eukaryotes with valine or alanine has been shown to be lethal (Hintermair et al., 2012; Hsin et al., 2011).

Dynamic glycosylation occurs at Thr4 (Comer and Hart, 2001; Kelly et al., 1993), Ser5 and Ser7 residues of the CTD by O-GlcNAc transferase (OGT) (Ranuncolo et al., 2012). Since glycosylation and phosphorylation are mutually exclusive, it suggests a functional role of Ser2/Ser5-G in preventing phosphorylation of CTD before its recruitment on promoters (Ranuncolo et al., 2012).

Two proline residues of the heptad (P3 and P6) can be found either in *cis* or in *trans* conformation, which further increases the combinations of CTD post-translational modification. Peptidyl-prolyl *cis/trans* isomerase (PPIases) is characterized to catalyze such conformational changes. Pin1 was found as a PPIase in human (Albert et al., 1999). It has been shown Pin1 activity depends on phosphorylation of its target site: Ser or Thr followed by Pro (Yaffe et al., 1997). A genetic study shows that Pin1 inhibits dephosphorylation of the CTD by Fcp1 (Xu et al., 2003) indicating that the proline isomerization status interplay with other CTD modification. In yeast, Ess1 is the homolog of Pin1. Like Pin1, it has been shown to interact with the CTD (Hanes et al., 1989; Morris et al., 1999) and isomerize the Ser5-Pro6 bond more effectively than the Ser2-Pro3 bond *in vitro* (Gemmill et al., 2005). It plays roles in RNA processing and termination (Bataille et al., 2012; Ma et al., 2012).

#### **1.4.4 CTD phosphorylation patterns within transcription cycle**

The cycle of eukaryotic transcription including initiation, elongation and termination stages is regulated at multiple steps. Within this process, dynamic phosphorylation patterns of CTD serine residues have been found to coordinate the recruitment of CTD binding factors to facilitate Pol II transcription (Hsin and Manley, 2012).

##### ***Transcription initiation***

RNA Pol II is recruited to the preinitiation complex (PIC) when CTD is hypophosphorylated (Lu et al., 1991). Following its binding to the promoter, Ser5 of yeast CTD is rapidly phosphorylated by either Kin28 (Cdk7 in human) (Feaver et al., 1991) or Cdk8 (Srb10 and Srb11 in human) (Liao et al., 1995).

These two kinase complexes are thought to have partially overlapping roles in transcription (Liu et al., 2004) to facilitate the dissociation of RNA Pol II from the PIC and elicit promoter escape (Sogaard and Svejstrup, 2007).

### ***Transcription elongation***

During transcription elongation CTD becomes further modified on Ser2 residues. CTDK-1 and Bur1/Bur2 complexes are considered equally responsible for Ser2 phosphorylation during Pol II elongation stage. However, Bur1/Bur2 complex is thought to contribute Ser2P at early elongation stage, while deletion of Ctk1 was shown to reduce Serine phosphorylation in the downstream region by 90% (Qiu et al., 2009).

CTDK-1 was the first CTD kinase (Lee and Greenleaf, 1991) discovered from yeast and its ortholog in human was later found to be CDK12 (Bartkowiak et al., 2010; Buratowski, 2009; Keogh et al., 2003). Ctk1 resides in CTDK-1 complex with the cyclin subunit Ctk2 and a unknown function subunit Ctk3 (Bartkowiak and Greenleaf, 2011). Cdk9 and its regulatory cyclin subunit form the positive transcription elongation factor b (P-TEFb) (Marshall et al., 1996). The closest relative to Cdk9 in *S. cerevisiae* is the Bur1/Bur2 kinase complex (Prelich and Winston, 1993).

### ***Transcription termination***

During transcription termination, CTD is de-phosphorylated by phosphatases to be recycled for the next round of transcription (Cho et al., 2001). Both Fcp1 and Ssu72 have been characterized as CTD phosphatase: Fcp1 is essential for cell survival in yeast and it converts Pol II from phosphorylated IIO form to unphosphorylated IIA (Archambault et al., 1997; Chambers and Dahmus, 1994). Ssu72 has been shown to preferentially remove Ser5P from the CTD (Hausmann et al., 2005), and its activity is enhanced by the peptidyl prolyl isomerases Ess1 (Krishnamurthy et al., 2009).

### **1.4.5 CTD interaction domain (CID)**

Like histone code, the combinatorial nature of CTD modifications leads to the hypothesis of a CTD code, where the patterns of modifications are read by the transcriptional machinery and other factors with various functional outputs. To date, a number of proteins and protein complexes have been reported to interact with the specific CTD modifications through distinct CTD interaction domains (CID).

Set2-Rpb1 interacting domain (SRI): Histone H3K36 methyltransferase Set2 and its orthologs

Set2/Hypb utilize conserved SRI domain to associate with phosphorylated CTD (Kizer et al., 2005; Sun et al., 2005). NMR studies show that both Set2 and Set2/Hypb SRI domain form three-helix left-handed bundles which are different from other CTD binding domains. The optimal CTD phosphorylation pattern for SRI interaction is found to be continuously phosphorylated Ser2/Ser5 within the CTD peptide containing two-heptad repeats and three flanking NH<sub>2</sub>-terminal residues (Li et al., 2005; Meinhart et al., 2005; Vojnic et al., 2006).

**WW Domain:** WW domain is characterized by a pair of conserved Tyr residues in -40 amino acid domain these domains specifically interact with Pro-rich peptides (Myers et al., 2001). Pin1 and Ess1 have been shown to interact with p-CTD through WW domain to determine the substrate specificity (Lu et al., 1999; Verdecia et al., 2000). Interestingly Set2 also contains WW domain which was thought to mediate p-CTD interaction (Krogan et al., 2003).

**Rtt103:** The transcription termination factor Rtt103 binds exclusively to di-heptad repeats phosphorylated on Ser2 through its N-terminal CID (Kim et al., 2004). Interestingly, the independent study shows that two CIDs of Rtt103 and the other termination factor, Pcf11, display cooperative binding features respectively on longer phosphorylated CTD (Lunde et al., 2010) indicating the potential interaction mode of these two factors at the 3' end of genes where Ser2 phosphorylation reaches highest density.

**FF domain :** FF domain is a 50-60 amino acid module containing conserved Phe residues near N- and C- termini. Human transcription factor TCERG1 /CA150, yeast splicing factor Prp40 and mammalian HYPA/ FBP11 interact with the phosphorylated CTD through multiple FF domains. (Carty et al., 2000; Morris and Greenleaf, 2000).

**Capping enzyme :** Cgt1 GTase subunit binds to a Ser5 phosphorylated CTD peptide containing four repeats through two CTD docking sites (CDS1 and CDS2) (Fabrega et al., 2004).

## 1.5 Summary

In this thesis, we focused on dissecting the molecular mechanism by which H3K36methylation is regulated in budding yeast. In the Chapter 2 we studied auto-regulation mechanism of Set2 and found a novel DNA interaction property of SRI domain which determines Set2 substrate specificity. Our biochemical study identified an auto-inhibition domain (AID) which resides in the middle portion of Set2 and demonstrated how Set2 activity is coordinated by AID and SRI domain to maintain the normal

H3K36methylation in cells. In Chapter 3, we demonstrated that phosphorylated Pol II CTD not only regulates the Set2 interaction but also affects Set2 activity *in vitro*. Our results suggest a new molecular model illustrating how Set2 methylates H3K36 when encounters elongating Pol II and transcribed nucleosome. Lastly, in the Chapter 4 we applied the knowledge learnt from the structural studies of other histone methyltransferases to indentify the novel F/Y switch residues in Set2 catalytic core and successfully obtained Set2 mutants with different product specificities. By using these mutants, we were able to dissect fuctions between H3K36me3 and lower methyl states in several pathways including cryptic transcription, histone chaperone, CTD proline isomerization and double-strand DNA repair. Our results suggest that the individual methyl-state of H3K36 plays non-redundant biological roles in cells.

The results from our studies add new insights into Set2-meidated H3K36 methylation. The knowledge we get from domain coordination, CTD regulation and product specificity control of yeast Set2 may help us understand more about functions of H3K36 methylation in regulating human developoment and tumorigenesis.

## **Chapter2 AN AUTOINHIBITORY MECHANISM DIRECTS THE HISTONE METHYLTRANSFERASE SET2 TO METHYLATE ONLY TRANSCRIBED CHROMATIN**

### **Abstract**

Set2-mediated H3K36 methylation ubiquitously functions in coding regions to control hypo-acetylated states, suppression of cryptic transcription and histone exchange, alternative splicing, DNA repair and recombination. It has been shown that H3K36me3 depends on the binding of Set2 to phosphorylated Pol II via its SRI domain. However, the detailed mechanism is poorly understood. Here we showed that fusion of the SET domain to Pol II was not sufficient for full activities of Set2, suggesting SRI may play additional regulatory roles to facilitate H3K36me3. We also found that the full length Set2 preferred nucleosomes as substrates while the SET domain alone preferred core histones. This substrate preference was partially determined by the DNA binding ability of SRI. In addition, we identified a novel auto-inhibitory role for the middle region of Set2 in regulating its catalytic activity. Remarkably, mutations at this region cause hyperactivities, which in turn lead to synthetic phenotype with an essential histone chaperone FACT. Our data suggest that the activity of Set2 is intrinsically regulated through multiple mechanisms and a temporal control for dynamic regulation of H3K36me is needed during transcription elongation process.

## Introduction

Nucleosomes are building blocks of eukaryotic chromatin that packages genomic DNA with histones. Each nucleosome consists of a core histone octamer wrapped by 146 base pairs of DNA. Histones are subjected to various post-translational modifications (PTMs) such as methylation, acetylation and phosphorylation. It was proposed that individual or combination of histone modifications mainly serve as a signaling platform for regulation of chromatin structure and almost all DNA-related processes (Jenuwein and Allis, 2001).

Histone H3 lysine 36 methylation (H3K36me) is an important histone “mark” that is ubiquitously present in all eukaryotes. In yeast, Set2 is responsible for all three states of methylation at H3K36, including K36me1 (mono-methylation), K36me2 (di-methylation), and K36me (tri-methylation) (Lee and Shilatifard, 2007; Strahl et al., 2002). H3K36 methylation is transcription-coupled and enriched at the middle and 3' end of RNA Pol II genes that are actively transcribed (Bannister and Kouzarides, 2011; Krogan et al., 2003). One of the key functions of H3K36me is the recruitment and activation of a histone deacetylase complex Rpd3S to transcribed regions (reviewed in (Yun et al., 2011)). Rpd3S deacetylates the nucleosome in the wake of Pol II passage, and this hypo-acetylated states of transcribed chromatin prevents spurious transcription being initiated from intragenic cryptic promoters (Carrozza et al., 2005b; Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2007b). Recently, H3K36me was also shown to suppress the exchange of newly synthesized histones at transcribed regions by decreasing the binding of H3 to histone chaperone Asf1 and recruiting the Isw1 chromatin remodeler (Smolle et al., 2012; Venkatesh et al., 2012). In higher eukaryotes, multiple SET domain containing methyltransferases have been identified to methylate H3K36 at different states (Wagner and Carpenter, 2012). Setd2, the homolog of yeast Set2, is the major H3K36 tri-methylase in *Drosophila* (Bell et al., 2007) and mammalian cells (Edmunds et al., 2008). Emerging evidence suggest that Setd2 plays critical roles in development (Hu et al., 2010) and tumor suppression (Newbold and Mokbel, 2010). Similarly, genetic alternations of other H3K36 methyltransferases, such as the NSD family members, were also linked to tumorigenesis of various cancers (Hudlebusch et al., 2011a; Hudlebusch et al., 2011b; Kassambara et al., 2009; Marango et al., 2008; Rosati et al., 2002; Wang et al., 2007). However the detailed causal relationship remains unclear. Lastly, Set2-mediated H3K36me also regulates DNA damage repair and genome stability (Fnu et al., 2011; Jha and Strahl, 2014; Pfister et al., 2014).

Set2-mediated H3K36 methylation is tightly regulated through multiple mechanisms. Set2 directly associates with elongating Pol II (Kizer et al., 2005; Krogan et al., 2003; Li et al., 2003a; Xiao

et al., 2003b). The c-terminal domain of the largest subunit of Pol II Rpb1 (CTD) is a large recruitment platform for many transcription related regulators (Corden, 2013b). The c-terminal Set2-Rpb1 interaction domain of Set2 (SRI) binds to CTD that is phosphorylated at Ser2 and Ser5 during elongation (Li et al., 2007a; Li et al., 2005; Vojnic et al., 2006). This association is crucial for the full activity and stability of Set2. It has been known that disruption of CTD kinases, such as CTDKI and the Bur1 complex causes the defects of H3K36me2 and H3K36me3 (Chu et al., 2007; Chu et al., 2006; Youdell et al., 2008). Similarly, mutation of the Paf1 complex, which also regulates the phosphorylation status of CTD, eliminates H3K36me3 but not H3K36me2 (Chu et al., 2007; Li et al., 2009a). Interestingly, expression of the SET domain alone can recover H3K36me2 to the wild type level with normal distribution pattern at transcribed regions (Youdell et al., 2008), suggesting that only H3K36me3 requires Set2 associating with Pol II. A recent study provided a plausible explanation for this methyl-state dependent requirement. Fuchs et al showed that Set2 has a short half-life which is controlled by proteasome-dependent pathway through the SRI domain (Fuchs et al., 2012). It was proposed that free Set2 (T1/2~ 9 mins) is capable of dimethylation of H3K36. However its association with the hyperphosphorylated Pol II not only stabilizes the protein but also increases its resident time at transcribed regions, which are essential for H3K36me3 (Fuchs et al., 2012). Moreover, Set2 also needs to make multiple contacts with nucleosomal substrates to achieve its full activity. Consistently, mutations of Set2 histone interaction motifs and histone surface residues lead to the comprised levels of H3K36me (Du and Briggs, 2010; Du et al., 2008; Psathas et al., 2009). Similarly, H3K36me3 is also influenced by the mutation of histone chaperones that modulate nucleosome assembly/disassembly, such as Asf1 (Lin et al., 2010) and FACT (Chu et al., 2006)

Set2 contains a highly conserved catalytic SET domain, WW domain, coil-coil domain and SRI domain. All Set2 homologs also have large middle region with unknown functions. Interestingly, a recent study reports that human Setd2 contains a proline-rich auto-inhibitory domain that inhibits the binding of the Set2 WW domain to human Huntingtin (Htt) protein (Gao et al., 2014). It was shown previously that Set1 contains a RNA-recognition motif (RRM) that is required for H3K4 trimethylation, while the RRM is controlled by an autoinhibitory domain (Schlichter and Cairns, 2005). In this study, we found that the middle domain of Set2 also functions as an auto-inhibitory domain (AID). We discovered that besides binding to Pol II, SRI is also responsible for DNA binding of Set2. AID mediates repression function through antagonizing the catalytic activity of SET and fine-tunes the both activities of SRI. We showed that Set2 AID mutants are hyper-active both *in vitro* and *in vivo* although H3K36

methylation levels in the mutants are also influenced by the stability of the mutant Set2. Interestingly, partial deletion of different regions of AID displays differential genetic interactions with histone chaperone FACT mutants suggesting the importance of a precise temporal control of K36me during dynamic elongation process.

## Results

### Engaging to phosphorylated CTD plays additional roles in Set2-mediated H3K36me3 beside stabilization and recruitment of Set2.

It was proposed that the binding of Set2 to phosphorylated Pol II stabilizes Set2 protein and increases its resident time at transcribed regions (Fuchs et al., 2012). An interesting study showed that fusing mRNA capping enzyme to the c-terminus of CTD can bypass the lethality caused by CTD S5A mutation (Schwer and Shuman, 2011). We therefore wondered if the artificial recruitment of Set2 to pol II CTD can bypass the requirement for SRI or CTD phosphorylation. To this end, we designed a series of mutants that contain several Set2 constructs fused to C-terminus of Rpb1 (**Fig 2.1 A**) and analyzed the levels of H3K36me in those mutants by western blotting. As shown in **Fig 2.1B**, fusing full length Set2 to Pol II did not enable Set2 to methylate H3K36 as efficiently as the wild type, as H3K36me3 level is reduced in the mutant (comparing Lane8 and Lane4). CTD-SET domain fusion displays no K36me3 and the wild type level of K36me2, which resembles the behaviors of the free SET domain (Youdell et al., 2008). Interestingly, when we fused Set2-618, which only lacks the SRI domain, to the Rpb1, we cannot detect any signals of K36me, suggesting that artificial recruitment of Set2 to Pol II is not sufficient for its activity, and SRI may play other roles. This result is consistent with previous finding that Set2-618 has no activity *in vivo* (Kizer et al., 2005).

To eliminate the possibility that fusion of Set2 to CTD disrupts its correct folding, we purified all three fusion Pol II using a Tandem Affinity Purification (TAP) method through TAP-tagged Rpb3. As shown in **Fig 2.1C**, silver staining showed that there is no endogenous Set2 co-purified with Pol II, which is consistent with previous reports (Li et al., 2003a). It was also noted that fusing Set2 to CTD did not alter the Pol II stability when the stoichiometry of the fusion Pol IIs were compared to that of wild type (Lane 5). We next tested the *in vitro* histone methyltransferase (HMT) activity of these CTD-fused Set2 which should all stem from the fused Set2. As shown in **Fig 2.1 D**, CTD-Set2-618 only showed background activity similar to what we observed *in vivo* (**Fig 2.1B**). Both SET and FL fusion display strong activities. We noticed that fused SET (CTD-Set2-261) appeared to be more active than FL. We speculate that in the FL fusion construct, the SET domain is situated in the middle of the protein. This may restrict its mobility which is important for catalysis.

We next sought to test if H3K36 methylation status at individual genomic loci follows the similar trends as we detected at the bulk levels. ChIP was performed to monitor the level of K36me at a model gene *STE11*. As shown in **Fig 2.1E** and **2.1F**, the methylation patterns at *STE11* in all three mutants were

generally consistent with those of bulk methylation. Specifically, H3K36me2 and H3K36me3 in both CTD-Set2-FL and wild type cells were distributed towards the 3' end of the ORF as previously reported (Li et al., 2007c). H3K36 methylation was almost abrogated in CTD-Set2-618 cells. CTD-Set2-261 could only achieve H3K36me2, suggesting that SET alone is not sufficient for H3K36me3 at any places even when it was tethered to Pol II with presumed longer resident time. Consistently, we showed that over-expression of Set2 SET still could not restore H3K36me3 (**Fig 2.2 A**). Deletion of transcription factor Paf1 complex subunit Cdc7 or Paf1 have been shown to cause CTD Ser2 phosphorylation defect and loss of Set2 protein levels (Fuchs et al., 2012). However we showed that restoration of Set2 levels in *CDC73* or *PAF1* null cells also led to the same phenotype (**Fig 2.2 B**), suggesting that the steady state levels of Set2 maintained by binding to phosphorylated Pol II CTD is not the only reason for the loss of H3K36me3.

Because CTD-Set2-FL did not di-methylate or tri-methylate the promoter even when it was forced to be recruited into these regions (A and B region), we wondered if the phosphorylation of CTD may be required for Set2 activity independent of recruitment. To test if CTD-Set2-FL activity still depends on CDTK-1 mediated Ser2P (Kim et al., 2010), we transformed a CTD-Set2-FL plasmid into  $\Delta$ *CTK1* cells. The endogenous Set2 did not interfere with this assay as it was degraded in the  $\Delta$ *CTK1* (Fuchs et al., 2012) and no K36me could be detected (**Fig 2.1G** lane 8). Interestingly, CTD-Set2-FL showed no K36me3 and markedly reduced K36me2 levels (**Fig 2.1G**). Decreased K36me was not due to the absence of Set2 because Set2 was fused to an essential Pol II Rpb1 subunit. Collectively, these data suggest that p-CTD/SRI engagement is essential for full activity of Set2, and this is potentially achieved by SRI looping back to CTD tail (**Fig 2.1H**). Intriguingly, at *STE11* promoter regions CTD-SET displayed more activity (**Fig 2.1 F**).

### **Full-length Set2 and Set2 SET domain display distinct substrate specificities.**

We next wanted to test if the enrichment of K36me2 detected at promoter by CTD-SET was due to different substrate specificities. Previous studies showed that Set2 activity requires Set2 interacting with histone H4 (Du et al., 2008). Therefore we first tested the HMT activity of SET and Set2-FL on core histones and nucleosomes. Two sets of substrates were used to compare the substrate specificity of SET and FL Set2: native Hela oligo-nucleosomes and core histones versus recombinant *Xenopus* nucleosomes and core histones. Different salt concentrations were also tested. As shown in **Fig 2.3A**, at high salt condition (200mM NaCl), SET and FL Set2 displayed similar activities on all the substrates. However at lower salt conditions (50 and 100 mM NaCl), SET domain preferred core histones, while FL

Set2 preferred nucleosomes. This finding was also confirmed by using H3/H4 tetramer as substrate (**Fig 2.3B**), in which FL Set2 and Set2 1-618 displayed much less activities than SET domain.

Set2 has been shown to prefer nucleosomal H3 (Strahl et al., 2002). Another Set2 family member NSD1, which perform mono- and di-methylated H3K36 in human is also shown to specifically target H3K36 with the presence of DNA (Li et al., 2009b). We hypothesized that the substrate preference of Set2 may be due to Set2/DNA contact. Multiple residues of Set2 were predicted to have potential DNA binding abilities and the majority of them were located around the SRI domain (**Fig 2.3 C**). Consistently, we observed that the Set2 activity depended on the linker DNA length (**Fig 2.3 D**). Compared with H3K36A nucleosome, Set2 is active on WT nucleosomes without any linker DNA (Yun et al., 2012) but showed stronger activity with nucleosome containing longer linker, whereas Dot1 worked equally on all nucleosome substrates. We then performed EMSA assay to test the DNA binding ability of Set2 (**Fig 2.3E**). The result showed that FL Set2 but not SET domain or 1-618 shifted double-strand DNA. Meanwhile GST-tagged SRI domain is sufficient to bind to DNA.

### **Set2 contains an auto-inhibitory domain for the H3K36 methyltransferase activity**

While FL Set2 and SET domain were active on respective H3 substrates, Set2-618 was always shown less active on either nucleosome (**Fig 2.1 B, D**) or histone H3/H4 tetramers (**Fig 2.3 B**). These results consist with previous *in vivo* studies that expressing SRI domain deleted Set2 in yeast caused almost total loss of H3K36me<sub>2/3</sub> (Fuchs et al., 2012; Youdell et al., 2008). Thus, we speculated that Set2 middle portion (aa262-618) inhibits Set2 activity. To test this hypothesis, we first expressed C-terminal truncated Set2 mutants in yeast strains and checked bulk levels of H3K36 methylation (**Fig 2.4 A**). The TAP tag sequence was integrated into the chromosomal copy of *SET2* to make SET2 1-261, 1-260 or 1-618 TAP strains. The result was consistent with previously published data that all the Set2 mutants were stably expressed whereas SET2 1-261-TAP and 1-260-TAP di-methylated H3K36 while SRI deletion (Set2 1-618-TAP) was shown to be inactive. Deletion of H3K36 demethylases slightly helped maintain more H3K36 methylation in cells.

We then purified wild type Set2 and Set2 aa1-618 from yeasts via their C-terminal TAP tags. The silver staining showed the similar result as previously published data that WT Set2 was co-purified with subunits of RNA Pol II complex (Li et al., 2003a) while no binding partner was found in Set2 aa1-618-TAP (**Fig 2.4 B**). The purified Set2 variants were then applied to *in vitro* HMT assay (**Fig 2.4 C**). Set2 1-618 TAP was shown to incorporate much less <sup>3</sup>H- labeled methyl group onto nucleosomes compared with Set2-Pol II complex. Combined with the result of bulk levels of H3K36 methylation in

cells, it suggests Set2 aa1-618 is a repressed form and the inhibitory effect likely comes from the region after SET domain (aa262-618). However, our *in vitro* result using native yeast Set2 1-618-TAP was not consistent with previously published data that recombinant FLAG tagged or CBP-tagged 1-618 is no less active than their FL counterparts (Du et al., 2008; Kizer et al., 2005). Since both of recombinant Set2 variants were purified from bacteria, we also purified GST-tagged Set2 from *E. coli* and found FL Set2 was highly degraded. The degradation happened presumably from its C-terminus because after GST elution we still got a lot of degraded contaminants (**Fig 2.4 E**). Thus we argued the discrepancies between previous data and our result is due to the degradation of FL Set2, which displays largely compromised activity.

To further confirm that Set2 1-618 is repressed, we purified FLAG-tagged FL Set2, SET domain (aa1-261) and SRI deletion (aa1-618) from baculovirus infected insect cells. The purified FLAG-tagged FL Set2 was much more homogeneous than the GST version from *E. coli* and all the Set2 variants that were purified from the same system also reached high level of purity (**Fig 2.4 E**). The HMT assay using either HeLa nucleosome or native yeast nucleosome as substrate showed that FL Set2 was more active than SET to perform H3K36me<sub>2/3</sub>. Meanwhile Set2 1-618 was least active and barely finished H3K36me<sub>2</sub> (**Fig2.4 F and G**). Thus we concluded that in Set2 1-618, the enzyme activity is repressed by an autoinhibitory domain (AID) which is located in the aa262-618.

### **The middle region of Set2 intrinsically suppresses the activity of the SET domain.**

We sought to map the minimally required inhibitory region by making a series of sequential C-terminal truncations. The secondary structure prediction suggests that there are five hypothetical alpha-helix motifs within aa262-476. In addition, to confirm that the inhibition of C-terminal truncation is not due to the structure rearrangement in SET domain, we introduced a tobacco etch virus (TEV) protease recognition site right next to the SET domain in all Set2 C-terminal truncation variants (**Fig 2.5 B**). The Set2 variant series includes FL Set2 and truncations at residue aa618, 476, 415, 338 and 302 respectively, each defining the last amino acid of the resulting mutant proteins. Moreover, we also expressed a SET-SRI fusion mutant in which the whole aa262-618 region was deleted while a TEV sequence and a poly-Glycine linker (four Glycines) were added right next to SET domain. All the purified FLAG-tagged Set2 mutants showed high level of purities by Coomassie blue staining (**Fig 2.5 C**). We also developed an experiment strategy to couple the TEV digestion reaction with HMT reaction to compare the activities of Set2 variants between protease and no protease treatment (**Fig 2.5 A**). As shown in **Fig 2.5 D** TEV protease completely released SET domain from its C-terminal portions except

FLAG-SET (aa1-261) which lacked a TEV site. Meanwhile, the FL Set2 was non-specifically digested by TEV protease (**Fig 2.5D lane1 and 2**). The HMT result of each individual Set2 variant activity between protease and no protease treatment shows that: 1. Without protease treatment, all the C-terminal truncations are less active than SET domain (**Fig 2.5E.black bars**). After incubating with protease, all the C-terminal mutants, except Set2 1-618 TEV, resume their activities to levels almost equal to SET domain (**Fig 2.5 E. grey bars**). 2. SET activity does not change upon protease treatment. 3. Interestingly, the FL-TEV and SET/SRI TEV Set2 are initially more active than SET domain but reduce their activities to SET domain levels after protease digestion. Our results strongly suggest that the activity change of Set2 mutants is not due to the mis-folding of SET domain upon C-terminus truncation because release of C-terminal portion of Set2 resumes their activities. Moreover, the SET is under regulations of two oppositely functioning domains: The middle region of Set2 (aa262-618) inhibits the SET domain activity while SRI domain antagonizes this inhibition and further stimulates it. These two opposing regulations happen *in cis* because the release of auxiliary domains leaves the SET domain working independently to display its intrinsic activity.

We have shown that SRI stimulates SET activity on nucleosomes via its binding to DNA (**Fig 2.3 E**). We wanted to further demonstrate the mechanism of inhibition from the middle domain of Set2. To do this, we used histone octamers as substrate which only contain each two of four types of histones to test the activities of Set2 variants (**Fig 2.5 F**). First of all, the fact that FL Set2 has less activity than SET domain to methylate histone octamers was addressed by earlier experiment. Secondly, SET-T-SRI which was super-active on nucleosome is now less active than SET domain. This result suggests SRI domain inhibits SET domain activity on octamer with the absence of the middle domain. Lastly but most importantly, the C-terminal truncated Set2 mutants displayed decreased activities while the whole middle domain was gradually restored. This result indicates that middle domain may directly inhibit SET domain activity on histone.

In summary, we have confirmed Set2 middle region (aa262-618) performs as an autoinhibitory domain (AID) to repress SET domain activity while SRI domain compensates AID inhibitory effect and further stimulates SET in free full length Set2 protein *in vitro*. Moreover, our octamer result suggests the inhibition from AID domain may be due to the repression of histone contact of SET domain.

### **AID interferes with the binding of SET to histone H3.**

We next exploit the mechanism by which AID inhibits SET activity. Set2 1-476TEV has been

shown to be as inactive as 1-618 TEV on both nucleosome and histone octamer (**Fig 2.5 E, F**) suggesting the region aa262-476 is sufficient to maximally inhibit SET activity. We wanted to test whether aa262-476 inhibits the interaction between SET domain and histones. FLAG beads pull-down assay was performed by using SET domain, 1-476TEV, and 1-618 TEV to test the binding of these proteins on H3/H4 tetramers (**Fig 2.6A**). The result supports our hypothesis by showing the 1-476TEV and 1-618 TEV proteins, which are immobilized by anti-FLAG resin, bind less histone H3 than SET domain does (**Fig 2.6 B and C**). For some reasons, histone H4 dissociates from H3 and is not pulled down by anti-FLAG resins. Our result shows that the AID domain inhibits the histone accessibility of SET domain by repressing the interaction between SET domain and histone H3.

### **AID plays important regulatory roles in Set2 activity *in vitro* and *in vivo*.**

We have demonstrated that aa262-476 is sufficient to inhibit SET domain but  $\Delta$ 262-476 disrupts the Set2/p-CTD binding. To further map the minimal deletion of AID domain that can disrupt the auto-inhibition, we deleted each individual or the combination of hypothetical alpha-helix motifs respectively (**Fig 2.7 A**). All the truncations were expressed in insect cells and purified out with high purity except the one with the third helix deleted ( $\Delta$  C,  $\Delta$  aa339-386). It is shown that this particular mutant is degraded (**Fig 2.7 B**) suggesting the region aa339-386 is important for protein integrity. The HMT assay result shows (**Fig 2.7 C**) that deletion of each helical-motif partially eliminates the AID inhibition and the deletion of all five helixes (aa262-476) maximally de-represses the activity indicating AID functions as a whole domain and each helical motif partially contributes to the inhibitory effect. Meanwhile the activity of helix B deletion (aa303-338) is most comparable to all five helix deletion among all the single helix mutants. Expressing each of the mutants in *SET2* null yeasts efficiently di-methylated H3K36 but H3K36me3 levels fluctuated (**Fig 2.7 D**). Only the deletion of helix region B (aa203-338) displayed increased H3K36me3 by 50% while others had compromised H3K36me3 (**Fig 2.7 E**).

We showed earlier that deletion of helix C caused the degradation of the mutant protein indicating AID domain is required for the protein stability in cells. Western blotting analysis also shows that the expression levels of FLAG tagged mutants are not equal to WT Set2 (**Fig 2.7 D**). We constructed the single helix deletion mutants to plasmids in which the ORFs are under the control of natural *SET2* promoter (**Fig 2.8 A**). Using antibody recognizing SET domain to detect the expression of individual mutant, we found all of the mutants except  $\Delta$ A were highly unstable compared with WT Set2 and none

of the mutants produced detectable levels of H3K36me3. In the other expression system in which *AHD1* promoter controlled ORFs were cloned in pRS415 backbone, the H3K36me3 can be detected in  $\Delta A$  and  $\Delta AB$  (**Fig 2.8 B**). We argue that, due to the disrupted protein stability by AID deletion, the steady state levels of mutants were compromised when under control of moderately over-expressed *ADHI* promoter. Thus the *in vivo* H3K36me3 levels fluctuated although the mutants were hyper active *in vitro*. We further compared the protein expression levels in all the plasmid systems we used and found that *GALI* promoter largely over-expressed Set2 in cells while *ADHI* promoter in pRS416 backbone had moderate expression levels. *ADHI* promoter in pRS415 backbone expressed as low levels of Set2 as natural *SET2* promoter did. This is probably because of the plasmid copy number issue. Importantly, the H3K36me3 levels correlated with the fluctuation of Set2 expression levels in different expression systems (**Fig 2.8 C**). These results partially explain the observation that H3K36me3 levels in some mutants are reduced.

Another possibility of the reduced *in vivo* H3K36me3 levels of hyper-active AID mutants is the loss of p-CTD recruitment by disrupting AID region. To address this, we performed p-CTD peptide binding assay. As shown in **Fig 2.7 F** the deletion of each individual helical motif had almost no effect on SRI/p-CTD binding while the mutant lacking the helix D gained moderate binding towards un-phosphorylated CTD peptide. However, the deletion of whole AID domain (SET-T-SRI) abolished p-CTD interaction suggesting SET domain inhibits SRI/p-CTD binding. We further purified GST tagged N-terminal truncations of Set2 containing SRI domain and different length of middle domain from *E. coli* to test their p-CTD binding abilities (**Fig 2.9 A**). The Coomassie blue staining showed GST FL Set2 and mutants were all under degradation (**Fig 2.9 B**). The intact FLSet2 specifically interact with phosphorylated CTD peptides (either on Ser2 or Ser5). SRI had similar binding affinity and specificity as FL Set2, suggesting SRI domain is sufficient for Set2/p-CTD interaction. Interestingly, while the other two N-terminal truncated mutants (aa262-733 and aa476-733) bound to phosphorylated CTD peptides, they also displayed different levels of binding affinity to un-phosphorylated CTD. Moreover, GST-tagged SET-T-SRI fusion protein showed significantly reduced phosphorylated CTD binding ability as its FLAG tagged counterpart did (**Fig 2.9 C**). Collectively, our CTD binding results indicate the p-CTD binding ability of Set2 SRI domain is also under regulations of middle domain and. The reason that hyper-active SET-T-SRI has reduced levels of H3K36me3 in cells is probably due to deficient p-CTD recruitment.

### **Mis-regulation of H3K36me leads to synthetic phenotypes between AID mutants and histone chaperone mutant.**

We checked H3K36 methylation status at *STE11* locus in yeast strains expressing each of three representative AID mutants: helix B deletion which displayed increased H3K36me<sub>3</sub>; helix C deletion which was highly unstable but still had WT levels of H3K36me<sub>3</sub>; and helix E deletion which showed decent protein levels but H3K36me<sub>3</sub> was significantly reduced (**Fig 2.7 D**). The result showed that in the cells expressing individual helix B or C deletion, the enrichment of H3K36me<sub>3</sub> was elevated at both promoter region and gene body of *STE11* locus while helix D deletion could not efficiently tri-methylate H3K36 (**Fig 2.10 A**). Meanwhile the helix C or D deletion caused elevated H3K36me<sub>2</sub> at both promoter and gene body of *STE11*. We also observed moderately increased acetylated H4 at 5' of *STE11* and this led us to question whether the Set2/Rpd3S pathway still functioned normally (Carrozza et al., 2005b). The Northern blotting analysis revealed mild cryptic transcription phenotype in cells expressing single helix or whole AID domain deletion mutants (**Fig 2.11 A**). The *STE11-HIS3* reporter strains displayed similar phenotypes in a more quantitative manner (**Fig 2.11 B**): the helix E deletion and the whole AID domain deletion had most severe cryptic transcription, while helix A and helix C deletion showed very mild phenotype. The severity among all the mutants correlated with the fluctuation of bulk levels of H3K36me<sub>3</sub> (**Fig 2.7 E**).

Previous study suggested that H3K36me<sub>2</sub> is sufficient for the Set2/Rpd3S pathway. However, we observed the cryptic transcription phenotype in our mutant strains where H3K36me<sub>2</sub> was intact (**Fig 2.7 D**). We speculate this is due to the altered Set2 dynamics. Disruption of AID domain interferes with Set2 stability (**Fig 2.8 C**) It is possible that the Set2 mutants quickly hyper-methylated H3K36 and got degraded. Thus mutants could not target transcribed nucleosome at the correct time when Rpd3S complex activity was required to be stimulated to deacetylate coding regions for the repression of cryptic transcriptions. Set2 has been shown to not be essential for cell survival but it genetically interacts with various transcription related factors (Biswas et al., 2006; Chu et al., 2006; Krogan et al., 2003; Ma et al., 2012). We asked how mis-regulated H3K36 methylation (e.g. excessive amount of H3K36me<sub>3</sub>, or mis-located H3K36me<sub>3</sub> in genome) affects normal cell function. Yeast cells expressing any of AID mutants grew normally (**Fig 2.10 B**) indicating the disruption of AID domain alone is not sufficient to influence cell growth. Previously published result showed that the growth defect by disrupting transcription dependent histone chaperon yFACT (facilitates chromatin transcription) complex (Spt16 or Pob3) (Belotserkovskaya et al., 2003) was rescued by deletion of *SET2* suggesting Set2 functions

oppositely to yFACT(Biswas et al., 2008b).

We expressed our AID mutants in *spt16-11/ΔSET2* yeast and grew them at permissive temperature (26°C) or non-permissive temperature (35°C). It was consistent with published results that *spt16-11/ΔSET2* transformed with empty vector grew faster than *spt16-11* strain in which endogenous Set2 was expressed, or *spt16-11/ΔSET2* strain expressing FL Set2 plasmid (**Fig 2.10 B**). Interestingly, the *spt16-11/ ΔSET2* strain expressing each the AID mutant displayed synthetic growth defect at 35°C. To this end, our result suggests that the AID domain is important to maintain normal Set2 function in Spt16 defect yeast strain at non-permissive temperature. Moreover, the deletion of *RCO1* which encodes the Rco1 subunit of Rpd3S complex totally rescued the synthetic growth phenotypes of AID mutants (**Fig 2.10 B**). This result is also in agreement with the previous finding that the genetic interaction between Set2 and Spt16 functions through Set2/Rpd3S pathway. Furthermore, the Set2 catalytic defect mutant C201A also rescued the phenotype of AID mutants confirmed that the normal Set2 activity is required for Spt16 pathway in cells.

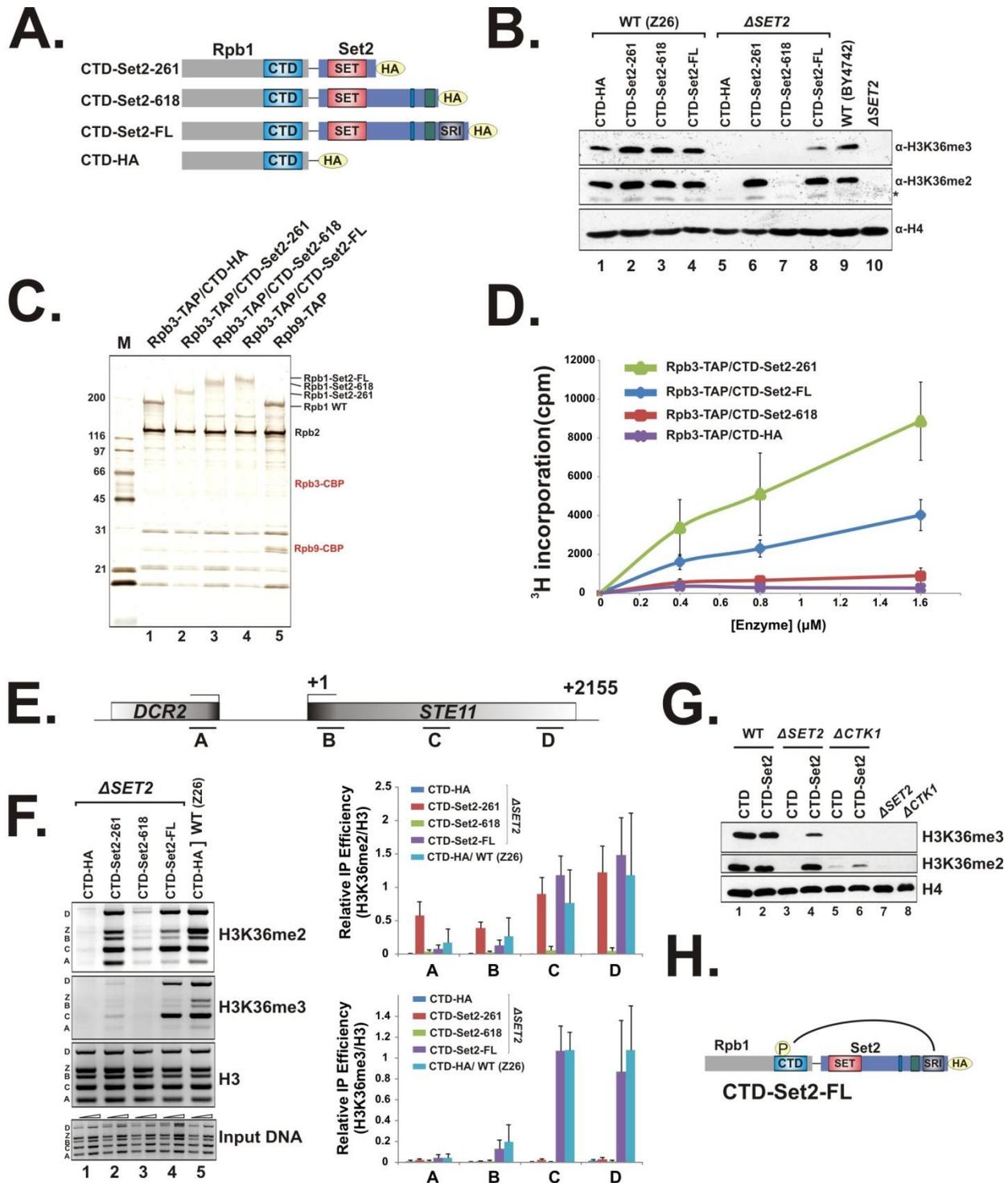
## Discussion

Phosphorylated Pol II CTD is required for tri-methylation at H3K36 and it has been proposed that p-CTD interaction helps maintain the methyltransferase Set2 stability (Fuchs et al., 2012). In this study we further demonstrate that Set2/p-CTD interaction plays additional roles in Set2 regulation other than simple association. This more intricate mechanism is different from that of other CTD interacting factors. For instance, in *S pombe*, interaction of mRNA capping enzyme with phosphorylated Ser5 on Pol II CTD is essential for its function. However, the lethality caused by S5A mutation can be bypassed by simply fusing this enzyme with CTD (Schwer and Shuman, 2011). Our results suggest that the additional function of SRI, besides interacting with CTD, may rest on its interaction with DNA (**Fig 2.3 E**). We showed that SRI regulated Set2/DNA contact to determine Set2 nucleosome preference and rendered it linker DNA length-dependent activity on reconstituted nucleosome (**Fig 2.3 A and D**). Meanwhile SRI antagonized the auto-inhibition effect from the middle region of Set2 (**Fig 2.5 E**).

In wild type cells, Set2 is recruited to elongating Pol II and localizes towards 3' end of ORFs (Pokholok et al., 2005). Interestingly, in our Rpb1 fusion system, both Set2-FL and SET domain were artificial tethered to Pol II. Therefore Set2 should be present along the entire gene. We found that CTD-Set2-FL methylate H3K36 in very similar pattern to wild type even though it was brought to the promoter and 5' regions of *STE11* (**Fig 2.1F**). H3K36 is acetylated predominantly at promoters of RNA polymerase II-transcribed genes (Morris et al., 2007). And local acetylation may inhibit Set2 from transferring methyl groups to the same H3K36. However, we did observe that ectopic H3K36me<sub>2</sub> occurred at this region suggesting at least some of the lysine 36 residues are available for di-methylation. Alternatively, CTD-Set2-FL could be repressed by rapid local histone exchange. It has been shown that during Pol II transcription, histones are highly dynamic (Li et al., 2007a) and the most rapid histone turnover occurs at promoters (Dion et al., 2007). We hypothesize that active histone exchange inhibit CTD-Set2-FL. Indeed, we revealed that Set2-FL and SET favored nucleosomes and core histones respectively (**Fig 2.3A**). We speculate that during transcription elongation at 5' of ORF, rapid exchange of histone leads to loss of nucleosome substrate for Set2-FL while provides more histones to SET. However at 3' end of the ORFs, the Pol II engaged Set2 has more chance to associate with nucleosomal histones through SRI/DNA interaction for its tri-methylase activity. We could not rule out other possibilities such as other modifications enriched at the promoter that specifically inhibit Set2-FL function.

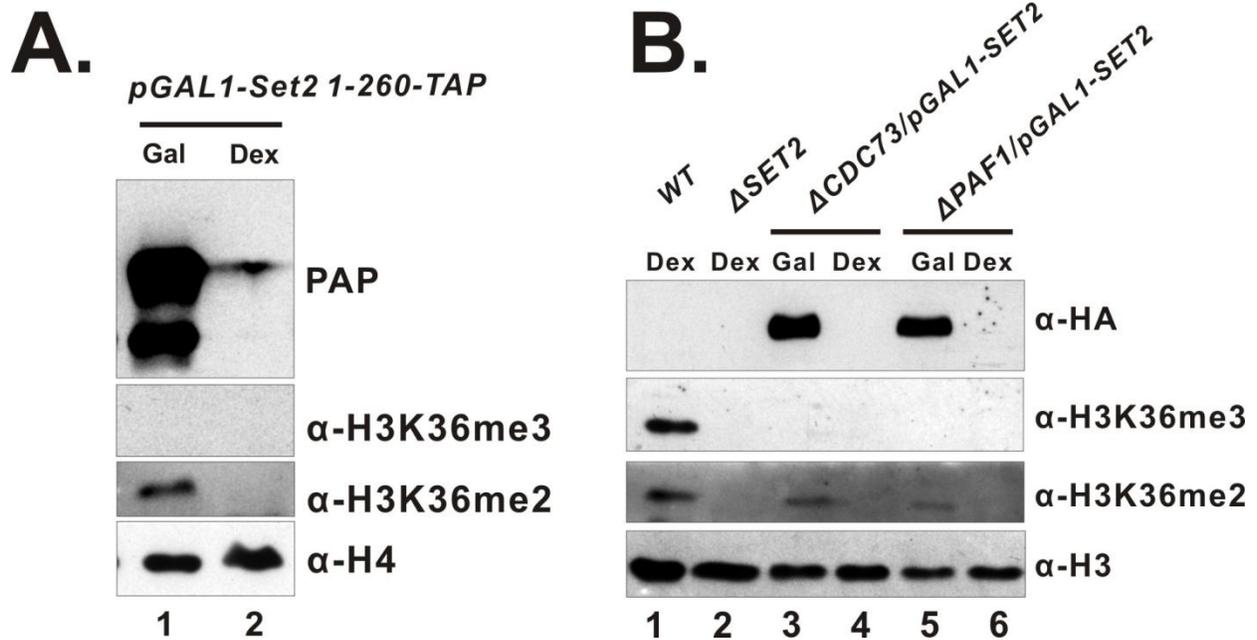
Auto-inhibition domain is a prevalent regulatory mechanism that provides tight “on-site”

repression to prevent uncontrolled activation and enable prompt response to cellular stimuli (Pufall and Graves, 2002). The auto-regulation also occurs in another SET containing protein Set1, a H3K4 methyltransferase, which has two unique domains to either inhibit or activate the enzymatic center (Schlichter and Cairns, 2005). Swi6, a heterochromatin protein 1 (HP1), in budding yeast, is also shown to release from auto-inhibited state to a hetero-chromatin spreading-competent state upon contacting H3K9 methyl mark and nucleosomal DNA (Canzio et al., 2013). In our study, we identified an autoinhibitory domain, which is antagonized by SRI domain to define the FL Set2 activity. Meanwhile the mis-regulated H3K36 methylation (e.g. excessive amount of H3K36me3) causes synthetic growth phenotype with histone chaperones (**Fig 2.10**) indicating it is important for auto-inhibition mechanism to fine-tune H3K36 methylation in cells. In human cells H3K36 methylation have been shown to be achieved by multiple SET domain containing methyltransferase belonging to the same family as Set2 (Dillon et al., 2005). In recently studies Set2/Hypb, the Set2 ortholog in human, has also been hypothesized as a tumor suppressor in breast cancer (Newbold and Mokbel, 2010) and sporadic clear renal cell carcinoma (Duns et al., 2010). Moreover, disruption of the Setd2-mediated H3K36me3 by somatic point mutation has been shown to link with leukemia (Zhu et al., 2014). Because of the high homology between Set2 and its ortholog in human (Sun et al., 2005), we speculate that the mechanism we found in yeast Set2 potentially helps us understand the regulation of H3K36 and provides epigenetic mechanism for tumorigenesis in human.



**Figure 2.1 Engaging to phosphorylated CTD plays additional roles in Set2-mediated H3K36me3 beside stabilization and recruitment of Set2.** (A) A Scheme for artificial recruitment of Set2 to Pol II by fusing indicated Set2 constructs to the c-terminus of Rpb1, the largest subunit of RNA polymerase II. (B) Fusing catalytic domain of Set2 to Pol II is not sufficient for H3K36me3. Plasmids including pWY043

(pYIA CTD-*SET2* FL), pWY044 (pYIA CTD-*SET2* 1-618), pWY045 (pYIA CTD-*SET2* 1-261) and pWY046 (pYIA CTD) were transformed into wild type (YYW120, Z26) and  $\Delta$ *SET2* (YYW121). Whole cell extracts were subjected to western blot using indicated antibodies. (C) Silver staining of TAP purified Pol II complexes. Indicated Pol II-Set2 fusion plasmids were introduced into the yeast YYW122 (a Pol II shuffle strain), which contains Rpb3-TAP for purification. (D) HMT assays using the PolII-Set2 fusion complexes as shown in (C) and HeLa oligonucleosome substrates. (E-F) H3K36 methylation status in Pol II-Set2 fusion strains was monitored by ChIP assays; (E) Illustration of PCR-amplicons around the *STE11* region (label as “A” to “D”) used in ChIP assays to monitor K36me; (F) Multiplex-PCR with indicated primer sets were performed to quantify the amount of immunoprecipitated DNA. The Z region (labeled as “Z”) located at a gene desert serves as internal control; however, we consistently detected H3K36me at this locus likely due to uncharacterized non-coding RNA transcription events. The relative IP efficiency is the ratio of IP efficiency of given antibodies at specific loci over the IP efficiency of histone H3. Data are represented as mean  $\pm$  SEM with at least three independent experiments. (G) CTD phosphorylation mediated by CTDK1 is required for H3K36me<sub>3</sub> even when full-length Set2 is artificially fused to Pol II. pWY046 and pWY043 were transformed into  $\Delta$ *ctk1* and control strains. Western blots of whole cell extracts of indicated strains. (F) A proposed model showing how Pol II-fused Set2 interacts with Rpb1 through phosphorylated CTD.

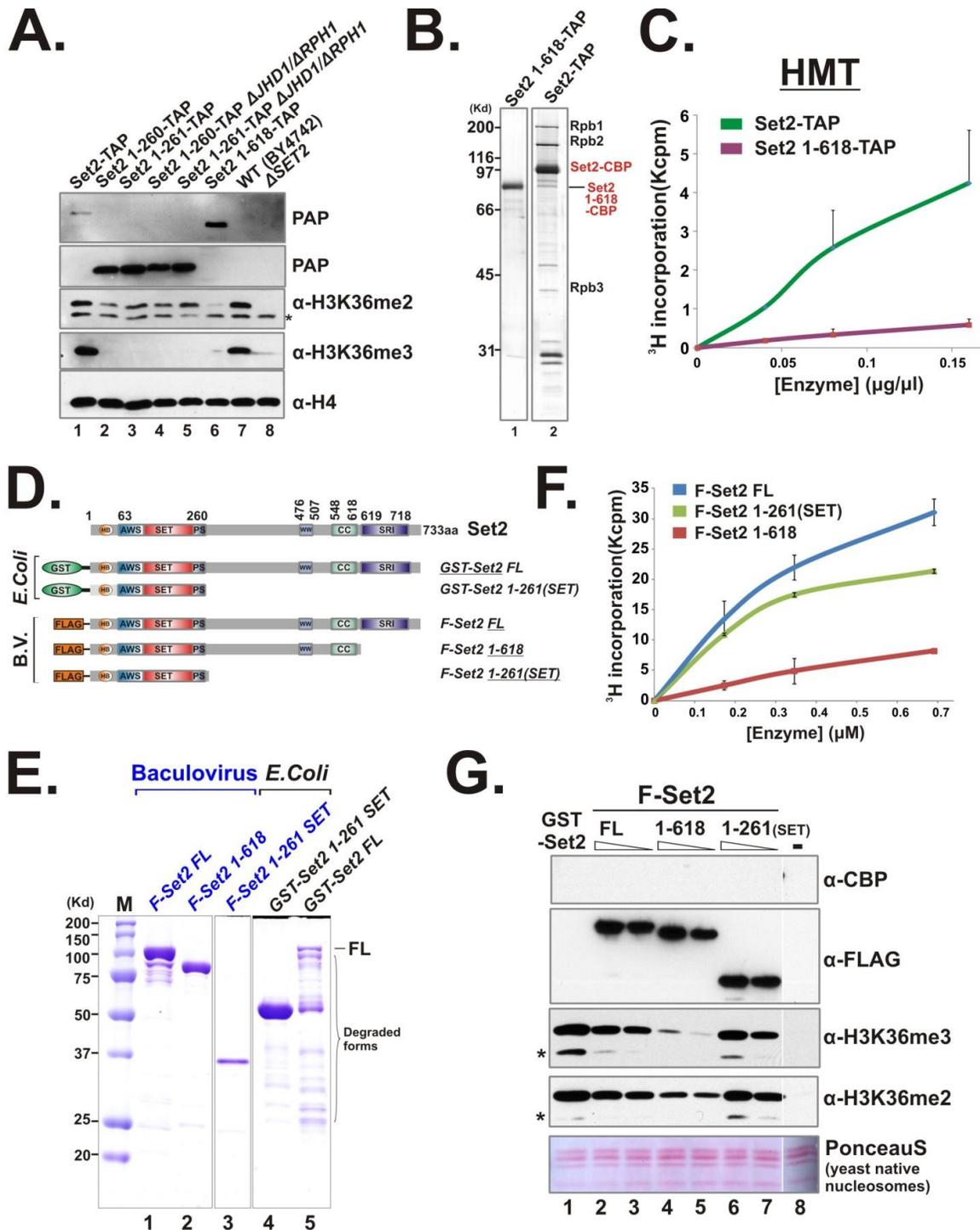


**Figure 2.2 Binding to phosphorylated Pol II CTD is essential for Set2-mediated H3K36me3.**

(A) Western blots of whole cell extracts from YYW042 (*HIS3::pGAL1-3HA-Set2 1-260-TAP::URA3*) grown in YP medium containing 2% galactose (Gal) or dextrose (Dex). (B) PAF.com is required for K36me3 even when Set2 is overexpressed. Western blots of cell extracts from YYW031 (*HIS3::pGAL1-3HA-Set2 /ΔACDC73::KAN*) and YYW032 (*HIS3::pGAL1-3HA-Set2 /ΔPAF1::KAN*) grown in YP medium containing galactose or dextrose.



and equal amount of each nucleosome/histone substrates. (B) HMT assay using *Xenopus* H3/H4 tetramers as substrates. (C) The residues of Set2 bind to DNA as predicted by the BindN+ (<http://bioinfo.ggc.org/bindn+/>) webserver. Predicted DNA binding residues are labeled with '+' and labeled in red, and non-binding residues are marked with '-', with confidence factors listed underneath. Key domains of Set2 were highlighted in different colors. (D). Set2 shows stronger activity on nucleosomal substrates containing linker DNA. Recombinant mono-nucleosomes with indicated DNA templates were used to test Flag-Set2-FL activity in standard HMT reactions. Recombinant GST-Dot1 was used as a control. (E) EMSA assay showing that SRI is responsible for the binding of Set2 to DNA. End-labeled 216bp DNA that contains a 601 positioning sequence was used. The truncated Eaf3 proteins were served as a positive control for proper DNA binding.

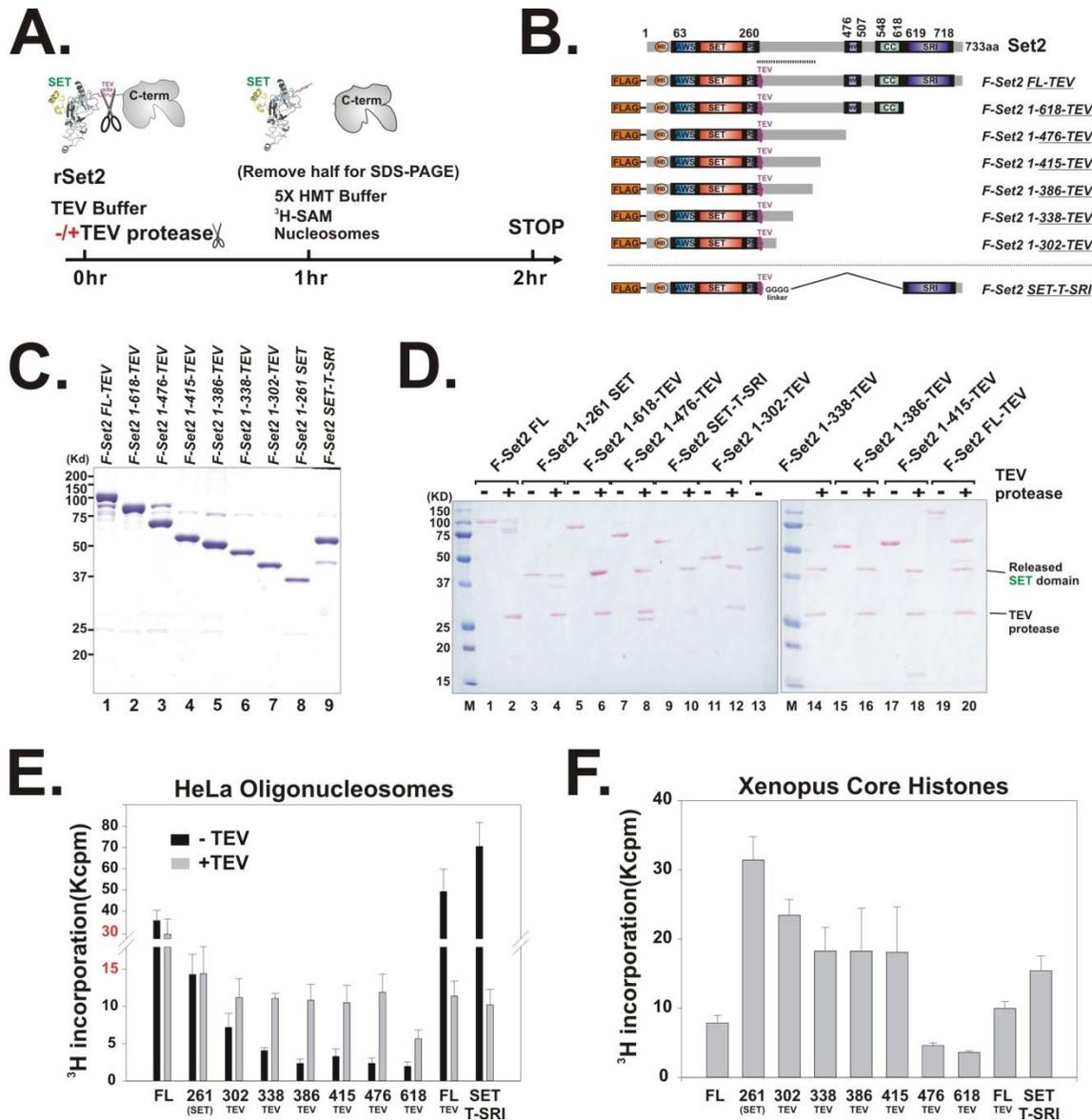


**Figure 2.4 Set2 contains an auto-inhibitory domain for the H3K36 methyltransferase activity**

(A) The Set2  $\Delta$ SRI mutant shows severely compromised H3K36 methylation *in vivo*.

Whole cell extracts from following strains: YYW087 (Set2-TAP:: URA3), YYW040 (Set2 1-618-TAP),

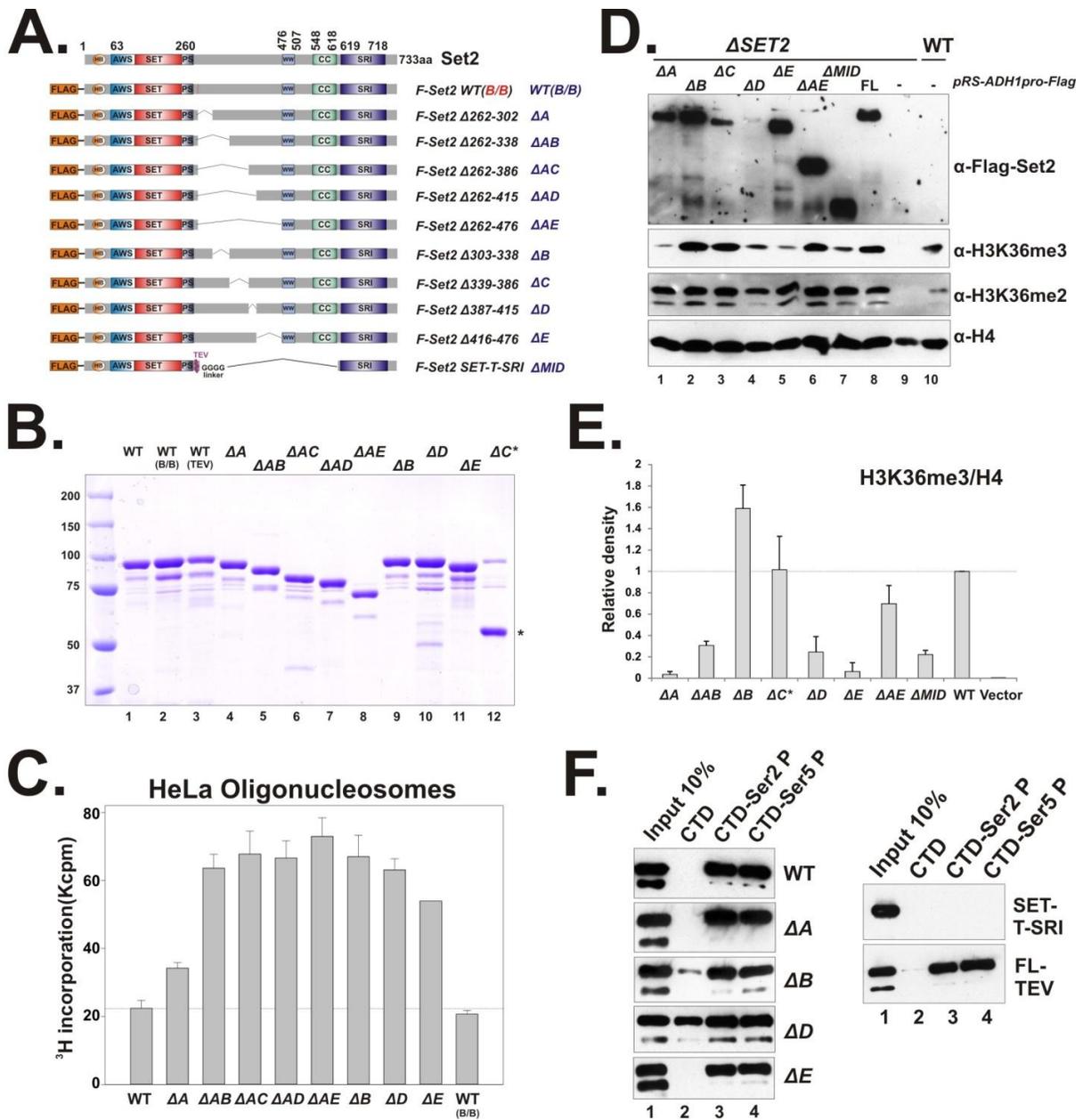
YYW044 (Set2 1-260-TAP), YYW045 (Set2 1-260-TAP $\Delta$ RPHI/ $\Delta$ JHD1), YYW127 (Set2 1-261-TAP) and YYW129 (Set2 1-261 TAP  $\Delta$ RPHI /  $\Delta$ JHD1) were subjected to western blotting with indicated antibodies. (B) Silver staining of TAP-purified Set2 WT and  $\Delta$ SRI complexes. (C) HMT assays using HeLa oligonucleosome substrates. HMT activity were represented as mean  $\pm$  SEM, n=3. (D) An illustration of recombinant Set2 and truncated derivatives: The SET domain, histone binding motif (HB), AWS domain, post SET domain (PS), WW domain, coiled-coil motif (CC) and Set2-Rpb1 interaction domain (SRI) are shown along with their position within full-length Set2. (E) Coomassie staining of purified Set2 and mutants as listed in (D) GST-Set2 purified from *E. Coli* shows severe degradation. (F) Standard HMT assay using HeLa oligonucleosome substrates. (G) HMT assay using native yeast nucleosomes that were extracted from  $\Delta$ SET2, which do not carry any H3K36me. The reaction products were detected using Western blotting with indicated antibodies. Ponceau S staining of yeast nucleosomes in each reaction served as loading controls.



**Figure 2.5** The middle region of *Set2* intrinsically suppresses the activity of the SET domain.

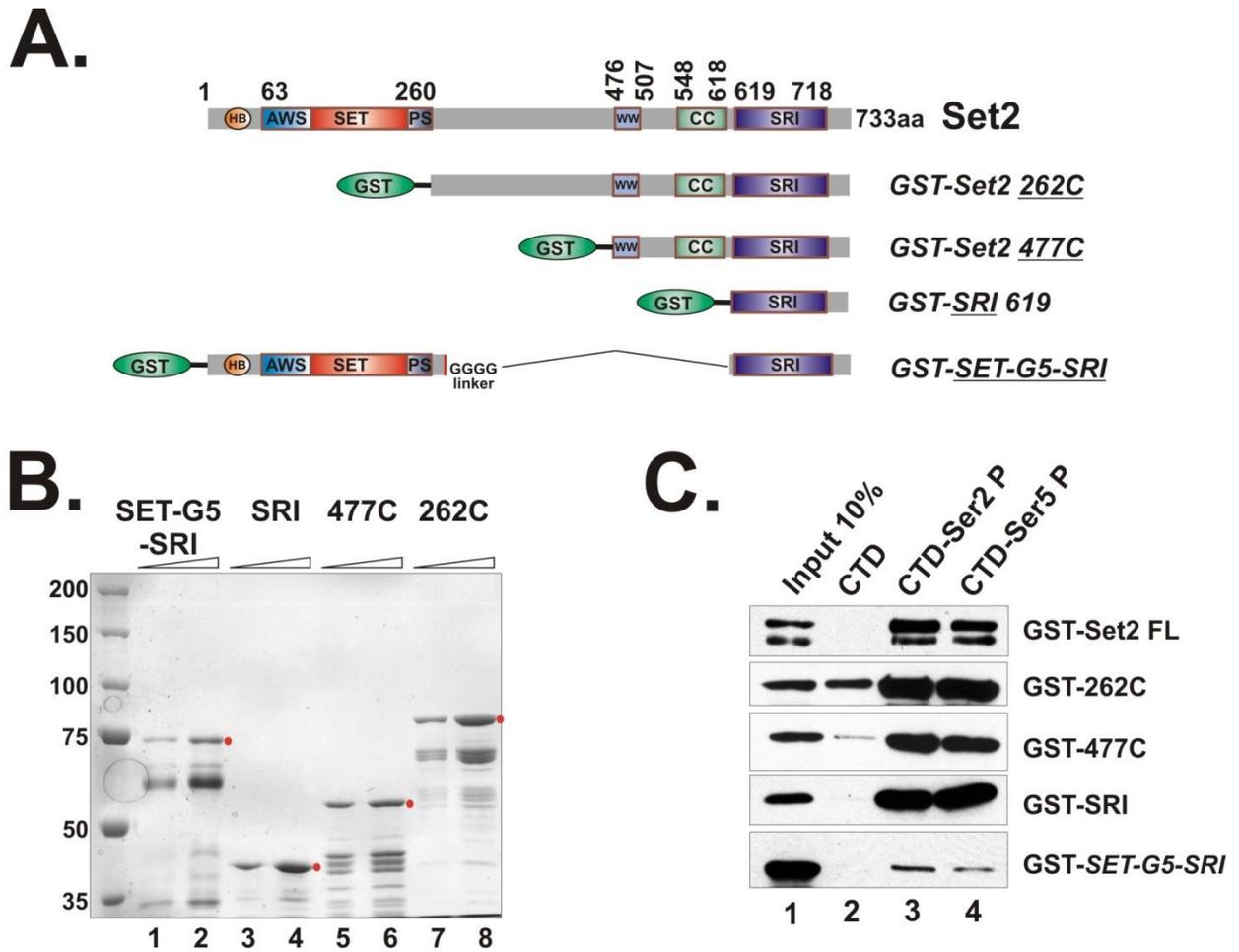
(A) An experiment scheme to test if SET activity is intrinsically inhibited. (B) A Schematic illustration of *Set2* and mutants in which a recognition site for TEV protease is engineered right after the SET domain. (C) Coomassie staining of *Set2* TEV proteins that were purified from insect cells. (D) Ponceau S staining showing the completion of TEV digestion. The bands represents TEV cleaved SET domain and TEV proteases were labeled. (E) Auto-inhibition domain of *Set2* only represses the SET activity *in cis*. HMT assays using HeLa nucleosome substrates to monitor the effect of TEV cleavage on *Set2* activity. Data are represented as mean  $\pm$  SEM. (F) HMT assay with indicated *Set2* derivatives and recombinant *Xenopus* core histone substrates





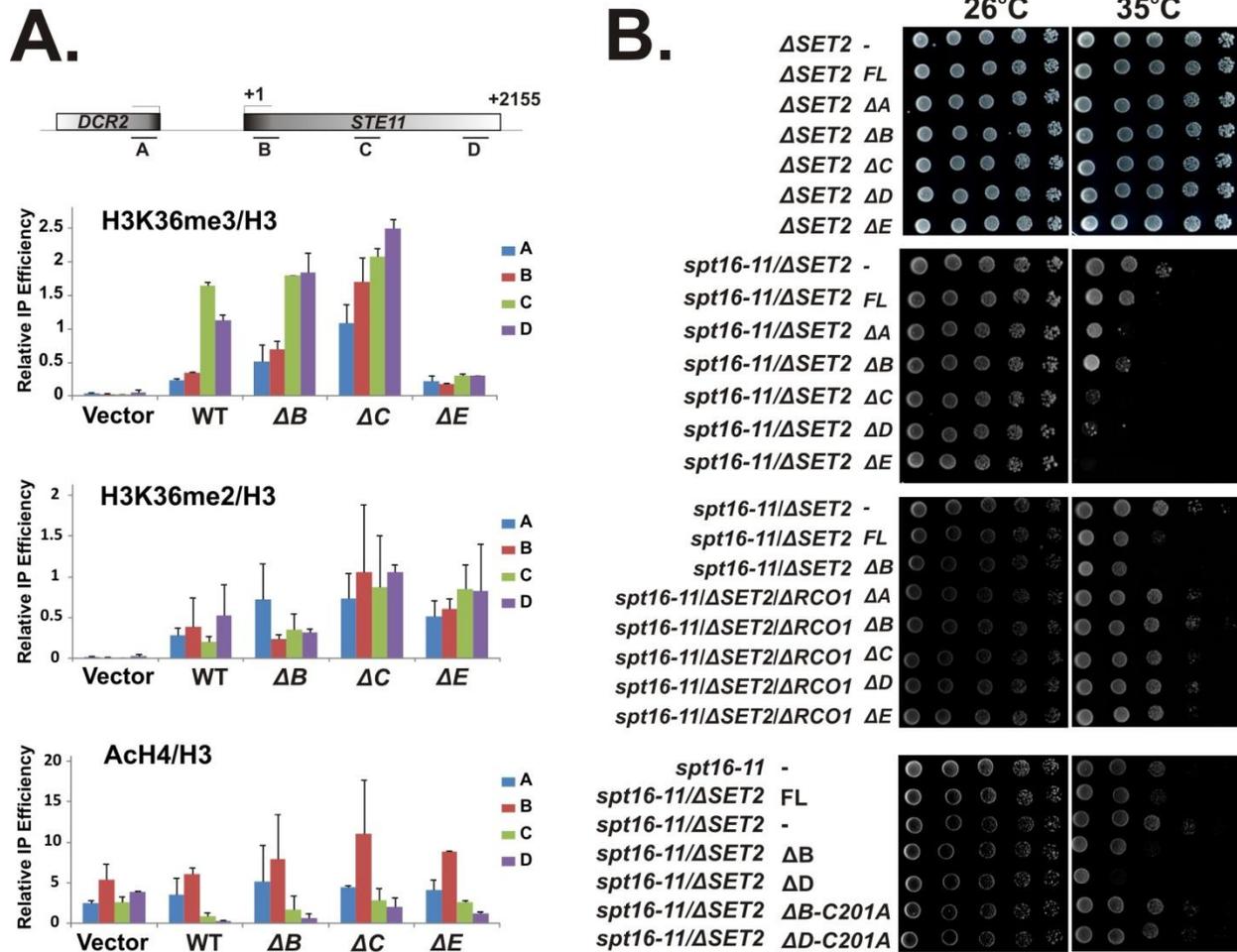
**Figure 2.7 AID plays important regulatory roles in Set2 activity *in vitro* and *in vivo*.** (A) Set2 constructs for mapping minimal deletion that causes de-repression of SET activity. (B) Coomassie staining of purified Set2 proteins. The asterisk indicates the degraded protein. (C) Integrity of the entire AID domain is required for controlling Set2 activity. HMT assay using HeLa oligonucleosome substrates. (D-E) Hyperactive Set2 mutants give rise to differential phenotype *in vivo*; (D) AID mutations make Set2 more sensitive to degradation, which in turn leads to distinct influences on K36methylation status. Western blot showing the level of tagged Set2 proteins and H3K36me levels in the mutants. (E) Quantification of H3K36me3 results shown in (D) based on three independent experiments. Data are represented as mean  $\pm$  SEM. (F) CTD peptide pull-down experiments showing that AID mutation did not compromise the binding of Set2 to phosphorylated Pol II CTD.



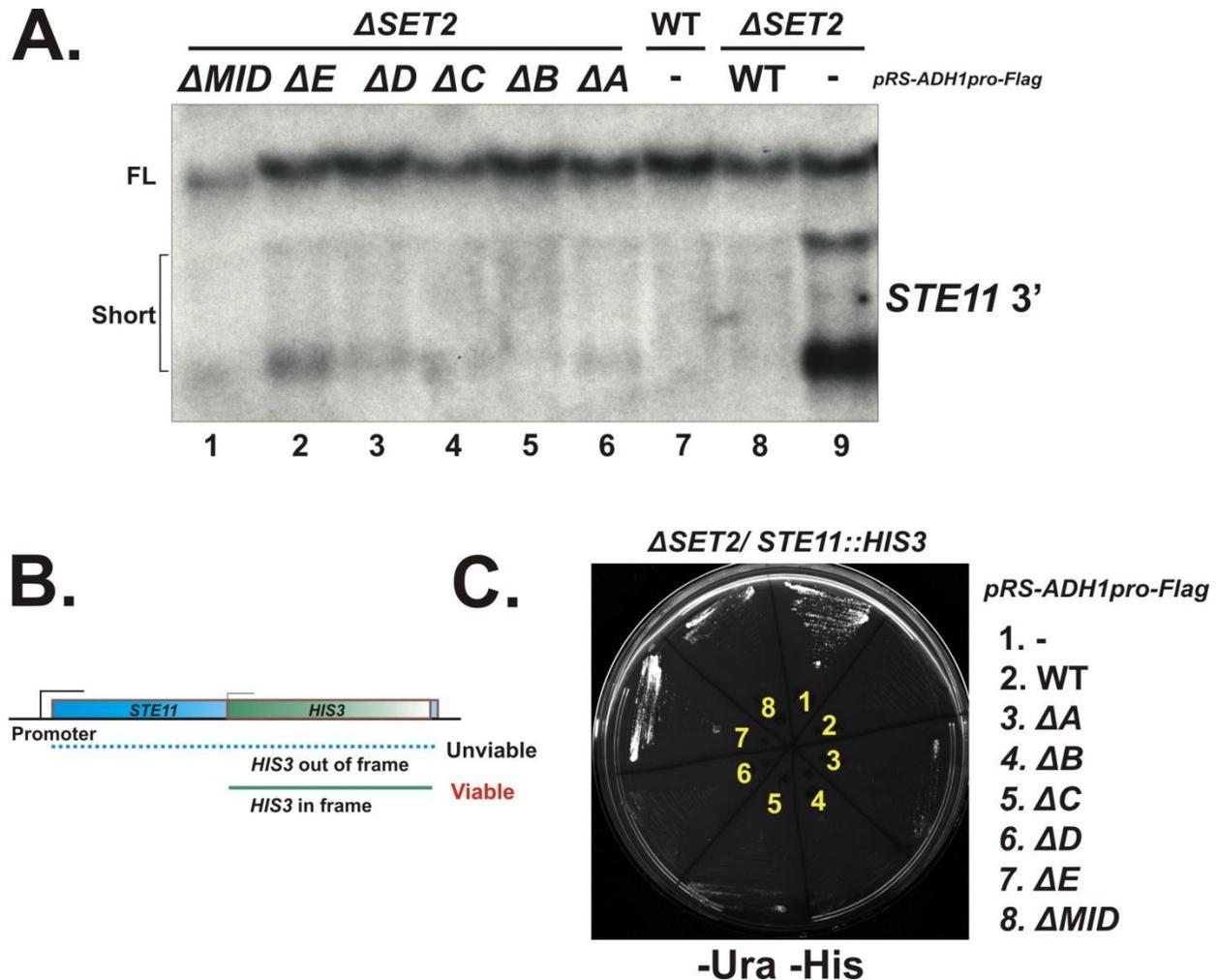


**Figure 2.9 AID regulates the binding of Set2 to phosphorylated CTD**

(A) Schematic showing GST versions of FL Set2 and its N-terminal truncations. The SET-G5-SRI was constructed by fusion of SET domain and SRI domain with five poly- Glycine linker. (B) The Coomassie staining of purified GST Set2 variants shown in A). (C) *In vitro* binding assay of recombinant GST Set2 variants and synthesized CTD peptides.



**Figure 2.10 Mis-regulation of H3K36me leads to synthetic phenotypes between AID mutants and histone chaperone mutant.** (A) Histone modification changes caused by AID mutations were determined using ChIP assays, as described in Figure 1E-F. IP efficiency were normalized to that of histone H3 and shown as mean  $\pm$  SEM,  $n=3$ . (B) AID mutants genetically interact with histone chaperone Spt16, which relies on their histone methyltransferase activities and the Set2-Rpd3S pathway. Each AID mutants were transformed into YYW010 ( $\Delta SET2$ ), YYW163 (*spt16-11/ΔSET2*) or YYW165 (*spt16-11/ΔSET2/ΔRCO1*). Transformants were grown at 26°C until saturation. 5 fold serial dilutions of each transformants were then spotted on two plates and grown at 26°C and 35°C (semi-permissive temperature). Photographs of the plates were taken after 2-3 days.



**Figure 2.11** *AID* mutations lead to cryptic transcription phenotype

(A) Total RNA from indicated yeast strains were subjected to Northern blots using *STE11*-3'RNA probe that was labeled through *in vitro* transcription. Full length and two short transcripts of *STE11* are indicated. (B) A diagram of cryptic transcription reporter strains, in which *HIS3* genes were integrated downstream of the *STE11* cryptic promoter (*STE11-HIS3*). The integration sites were selected such that the *HIS3* gene is out of frame with relation to the *STE11* coding region, and the functional His3 can only be produced when the *HIS3* transcript initiates at the cryptic promoter of *STE11* (Ruan and Li, unpublished). This strategy was first developed by the Winston lab (Cheung et al., 2008). (C) Indicated plasmids were transformed into YCR377 (*STE11-HIS3/ΔSET2*). The resulting strains were grown on SD-URA/-HIS plate at 30°C for 5 days.

## Materials and Methods

### *Construction of Plasmids and Yeast Strains:*

Plasmids and yeast strains were constructed through standard procedures and listed in Table 2.1 and 2.2 respectively. To engineer a TEV cleavage site into Set2, two fragments of DNA were prepared separately as illustrated in Figure 4B. One fragment contains the SET domain of Set2 flanked by the 5' Xho1 site and the 3' end BamH I site plus the TEV-recognition site and a flexible linker. The other one includes the rest of C-terminal part of wild type SET2 (or different truncations) with 5' BamH I site and 3' Not1 site. Those two fragments were joined through BamHI site and subcloned into Xho I/Not I sites of pBL532 through three-way ligation. To generate deletion mutants at the middle of Set2 (Figure 6A), similar approaches were taken. In this case, the 5' fragment contains the sequence corresponding to the N-term pieces before the break with a 3' BamHI site (without linker), while the 3' fragment covers the C-terminal part of Set2 with a 5' BamHI site. To construct an expression plasmid for *SET2* under the control of *ADHI* promoter, pWY080 (a gift from Dr. Brian Strahl) was modified by inserting a short adaptor containing FLAG tag sequence and Xho I/Not I restriction sites, thus allowing subcloning all Set2 mutants fragments from the baculovirus system to the yeast vectors through the same Xho I/Not I sites. To construct Pol II-Set2 fusion plasmids, an intermediate vector, pBL875 which covers the sequence between natural Kpn I and SnaB I of *RPB1*, was created. Two restriction sites (Sma I and Xho I) and the sequence for HA tag, in this order, were inserted between the last codon of heptad repeat sequence and the rest of CTD “fluff” of *RPB1* through three-way ligation. The coding sequences for *SET2* and its derivatives were then subcloned into the Xho1 site of pBL875 to form pWY041 (FL), pWY042 (1-618) and pBL876 (SET domain). The Kpn I/SnaB I fragments from these intermediate vectors were subsequently cloned into the Kpn I/SnaB I sites of pBL855 (pY1A) to generate pWY043 (CTD-Set2 FL), pWY044 (CTD-Set2 1-618), pWY045 (CTD-Set2 SET) and pWY046 (CTD-HA, from the empty vector of pBL875). The resulting constructs were shuffled into YYW120 (Z26) and YYW121 (Z26,  $\Delta SET2$ ) using the standard 5'-FOA method so that Pol II-Set2 fusion is the only source of RNA polymerase II in final strains.

To make YYW031(*HIS3::pGAL1-3HA-Set2 / $\Delta CDC73::KAN$* ) and YYW032 (*HIS3::pGAL1-3HA-Set2 / $\Delta PAF1::KAN$* ) strains, YYW026-A(*HIS3::pGAL1-3HA-Set2, MATa*) was mated with YBL703 ( *$\Delta CDC73::KAN$  MAT alpha*) and YBL 701 ( *$\Delta PAF1::KAN$  MAT alpha*) respectively. Selected diploid cells were seeded in sporulation media containing 1% potassium acetate for about 1

week due to inefficient sporulation of the mutants. Spores were dissected and the presence of *HIS* and *KANMX* cassettes was screened by PCR. Lastly, western blotting with anti-HA antibody was performed to confirm proper expression of HA-tagged Set2 in Dextrose and Galactose containing media.

***Preparation of whole-cell extracts for immunoblotting:***

Cells were grown in 3ml YPD or synthetic drop-out media supplemented with 2% dextrose or 2% Galactose at 30°C overnight until OD reached 1-1.2. Cell pellets were re-suspended in 45µl of STE buffer (containing 500mM NaCl, 10mM Tris HCl PH8.0 and 1mM EDTA) and 40µl of 3xSDS buffer. 100µl of 0.5mm glass beads were then added and boiled at 95°C for 5 minutes. Samples were then vigorously vortexed to break the cells and cell suspensions were clarified by spinning at 14000rpm for 5mins. Anti-H3K36me3 (Abcam, 9050), anti-H3K36me2 (Abcam, 9049), anti-H4 (Abcam, 10158) and anti-FLAG HRP (Sigma, A8592) antibodies were used according to manufacturers' suggestion. For anti-H3K36me2 antibody, 5% BSA was used for blocking and 1% BSA were present in all antibody incubation.

***Protein Purification:***

*Insect cell system:* 100ml Sf21 cells ( $1 \times 10^6$  cells/ml) supplemented with 10% FBS (Sigma), 1% Penicillin Streptomycin (Life Technologies) were infected with proximally 10ml of each P2 virus and incubated at 27°C for 48 hours. Cells were collected and washed with 10ml 1X PBS buffer. Cell pellets were then lysed in 10ml BV-lysis buffer (50mM HEPES PH7.9, 500mM NaCl, 10% glycerol, 0.5mM EDTA, 2mM MgCl<sub>2</sub> and 0.2% Triton X-100) on ice for 30 minutes. The lysates were clarified by ultracentrifugation (Beckman 50.2 Ti 40,000 rpm) for 30 minutes. 200 µl of Anti-Flag M2 affinity beads (Sigma) was added to the supernatants and incubated at 4 °C for at least 2 hours. Beads were then washed with 10ml BV-lysis buffer for 3 times and eluted with 600µl of 500µg/ml of 3xFLAG peptides (Sigma F4799) twice. The eluents were concentrated to about 150µl using 10kd cut-off concentrators (Amicon).

*TAP purification from yeast:* The TAP purification was performed essentially as described in previous publication (Li et al., 2003b). Briefly, 6 liters of TAP-tagged Set2 derivative strains were grown in YPAD media supplemented with 2% Dextrose at 30°C OD<sub>600</sub>=2-3. The cell pellets were re-suspended in an Extraction buffer (E buffer) (40mM HEPES pH 7.5; 350mM NaCl; 10% Glycerol; 0.1% Tween 20, supplemented with Protease Inhibitor Cocktail (Sigma P8215 ) and disrupted through bead-beating

(Biospec). Crude whole cell extracts were clarified through ultracentrifugation (Beckman 50.2 Ti at 45,000 rpm) and then incubated with 600 $\mu$ l IgG-sepharose (GE) at 4 °C overnight. Bound proteins were cleaved from the beads by 100 units of TEV protease (Invitrogen) at 18 °C for 3hours. The eluent were mixed with 3ml of Calmodulin Binding Buffer (10mM Tris.HCl pH8.0, 150mM NaCl, 1mM Magnesium acetate, 1mM Imidazole, 2mM CaCl<sub>2</sub>, 10mM  $\beta$ -ME, 0.1% NP40, 10% Glycerol) and 3 $\mu$ l of 1M CaCl<sub>2</sub>, and subsequently incubated with 600 $\mu$ l Calmodulin-Sepharose (GE) at 4 °C for 1hour. The bead bound proteins were finally eluted by Calmodulin Elution Buffer (10mM Tris.HCl pH8.0, 150mM NaCl, 1mM Magnesium acetate, 1mM Imidazole, 2mM EGTA pH8.0, 10mM  $\beta$ -ME, 0.1% NP40, 10% Glycerol) and concentrated using Amicon concentrators.

### ***Recombinant Histone Purification and Nucleosome Reconstitution***

Xenopus recombinant histones (H3, H4, H2A, H2B) were individually expressed in BL21CodonPlus-RIL (Stratagene) cells and purified as described (Li et al., 2007b). Histone octamers were assembled and fractionated through gel-filtration column Superdex 200. To generate mono-nucleosomes with different linker DNA lengths, plasmids that carry 16 copies of 216-bp, 153-bp or 147-bp DNA fragments were digested by EcoRV and the template DNA were purified using the 491 Prep-cell (Bio-rad) (Yun et al., 2012). These DNA fragments were then mixed with histone octamers at a pre-determined ratio in high salt buffer, and dialyze against a serial of buffer with decreasing salt concentration (Yun et al., 2011). The reconstituted products were finally purified through a 4% native gel using 491 Prep-cell (Bio-rad), and concentrated to about 1 $\mu$ g/ $\mu$ l.

### ***Yeast native nucleosome purification***

Yeast nuclei were prepared as previously described with minor modifications (Li and Reese, 2001). Briefly, 3 liters of YYW010 ( $\Delta$ SET2) cells were cultured to OD 600=1.0. Cell pellets were re-suspended in cold SB buffer (40mM HEPES.NaOH pH7.5, 0.5mM MgCl<sub>2</sub> and 1.4M sorbitol and 10mM beta-ME/1Mm PMSF) and digested with Zymolyase (Seikagaku 100T) up to 30mins at 30°C. The spheroplasts was washed with SB buffer twice and resuspended in cold FB buffer (20mM PIPES.NaOH pH6.5, 0.5mM MgCl<sub>2</sub>, 18% Ficoll 400) and homogenized using a Thomas Pestle Tissue Grinder (Size C). The resulting suspension was laid over 10ml GB buffer (20mM PIPES.NaOH pH6.5, 0.5mM MgCl<sub>2</sub>, 7% Ficoll 400, 20% Glycerol) and centrifuged at 11,500rpm using a HB6 rotor. The pellets were resuspended in 20ml of FB buffer, and spun at 4,500rpm to remove cell debris. The nuclei were finally collected by centrifugation at 11,500rpm for 30mins, and extracted three times with 120ml of EBX

buffer (50mM HEPES.NaOH pH7.5, 2.5mM MgCl<sub>2</sub> 400mM KCl 0.25% Triton X-100). The remaining chromatin pellets were resuspended in 10 ml MNase digestion buffer (10mM HEPES.NaOH pH7.5, 0.5mM MgCl<sub>2</sub> and 0.05 CaCl<sub>2</sub>) and sonicated briefly to break the clumps. After a small scale MNase titration test, the optimal concentration of Micrococcal nuclease (MNase) was applied to the rest of chromatin suspension at 30°C for 15mins. MNase digestion was stopped by addition of EDTA. The digested oligo nucleosomes were then fractionated on a Sepharose CL-6B column using 0.1M HB buffer (20 mM HEPES.NaOH pH7.5, 1 mM EDTA, 100 mM NaCl, 10 % Glycerol, 1 mM βME). Nucleosomes were pooled based on the size of DNA, and concentrated using 30kd cutoff concentrators (Amicon).

### ***Histone methyltransferase assays (HMT)***

Standard HMT reactions were carried out in a 20μl system using 1xHMT buffer (50 mM Tris.HCl pH 8.0, 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM DTT, 5 % Glycerol) supplemented with either 0.5μl of radio-labeled <sup>3</sup>H S-Adenosyl methionine (80 Ci/mmol, Perkin Elmer) or 10μM cold S-Adenosyl methionine (Sigma) (Li et al., 2003b). Scintillation counting-based filter-binding assay was performed as described previously (Eberharter et al., 1998). To detect specific methylation states on given histone substrates, western blots were performed using antibodies against H3K36me<sub>3</sub> (Abcam 9050) and H3K36Me<sub>2</sub> (Abcam 9049). As for the TEV-digestion HMT experiments described in Figure 4, WT Set2 and its derivatives that each contains an internal TEV cleavage site were incubated with 2.5units of TEV protease (Life technologies) in 10μl of TEV Buffer (10mM Tris.HCl pH=8.0, 0.1% NP40 and 1mM DTT) at 30°C for 1 hour. Half of the reactions were directly mixed with 3X SDS loading buffer for SDS-PAGE electrophoresis. 15μl of the reaction master mix (for each reaction: 4μl of 5x HMT Buffer, 0.5μl of <sup>3</sup>H S-Adenosyl methionine, 1μg of HeLa oligonucleosomes and 9.8μl H<sub>2</sub>O ) was then added to the other half of each reaction and incubated at 30°C for 1 hour. <sup>3</sup>H-incorporation was finally measured via filter-binding assays. For salt-titration experiments , 5xHMT-0-NaCl buffer (250 mM Tris.HCl pH 8.0, 5 mM MgCl<sub>2</sub>, 10 mM DTT, 25% Glycerol) was used and each salt condition was taken into consideration of the NaCl concentration from each substrates (HeLa nucleosomes 0.6M; Hela core histone 2.5M, recombinant nucleosomes 0M and recombinant histone octamers/tetramers 2M) and adjusted by addition of 5M NaCl.

### ***Chromatin immunoprecipitation (ChIP)***

200ml of each yeast culture were grown to reach OD<sub>600</sub>=1.0. 1% formaldehyde was used for

crosslinking at RT for 15mins. Standard ChIP procedures were described previously (Li et al., 2007b). 1 $\mu$ l of anti-H3 (abcam 1709), 0.5 $\mu$ l of anti-trimethylated H3K36 (abcam 9050), 1.5 $\mu$ l of anti-dimethylated H3K36 (abcam 9049) were used for immunoprecipitation. Input DNA and immunoprecipitated DNA were quantified using multiplex PCR. Primers mixture contains 5 primer pairs corresponded to: -879 and -745 of *STE11* promoter (A)(P1397/1398), 87 and 330 of *STE11* 5' (B)(P1399/1400), 925 and 1111 of *STE11* coding region (C)(P1401/1402), 1839 and 2129 of *STE11* 3' (D)(P1403/1404), 4193 and 4477 of ORF-free region of ChrI (Z) (P1740/1741). Band intensity was quantified using the UN-SCAN-IT V6.1 software (Silk Scientific, Inc.)

### ***CTD peptide pull-down Assay***

Biotinylated peptides that corresponding 4 heptad repeats of unphosphorylated (CTD), or Serine2 phosphorylated (Ser2P) or Serine 5 phosphorylated (Ser5P) of Pol II CTD were immobilized on Dynabeads® M-280 Streptavidin (Invitrogen) at 4 °C for 2 hours. The beads carrying 8  $\mu$ g of each peptide were then used for each pull-down reaction with about 1 $\mu$ g full-length Set2 or equal molar amount of mutant proteins in Peptide Binding Buffer (25mM Tris.HCl pH 8.0, 50mM NaCl, 5% Glycerol, 0.03% NP40 and 1mM DTT). The bound proteins were detected through western blots using anti-Flag HRP (Sigma) or anti-GST (Santa Cruz) antibodies.

### ***Northern blotting assay***

Total RNA were extracted by vortexing cells in 200 $\mu$ l of 0.5mm glass beads(Biospec), 320 $\mu$ l of RNA Prep Buffer (0.5M NaCl, 0.1M Tris-HCl pH 7.5, 10mM EDTA and 1% SDS) and 250 $\mu$ l of Phenol-Chloroform Isoamyle-alcohol (Fisher). RNAs were then Ethanol precipitated, resolved on 1% agarose-formaldehyde gels and transferred onto Zeta-probe membrane (Bio-Rad). UV-crossed membranes were subjected to hybridization in Denhardt's based buffer (6x SSC, 5X Denhardt's solution, 0.5% SDS and 0.1 mg/ml of sonicated salmon sperm DNA). 3' *STE11* probe was prepared through *in vitro* transcription reaction using T7 RNA polymerase (MEGAscript® Kit, Life Technologies Corporation) and 5 $\mu$ l of  $\alpha$ -<sup>32</sup>P labeled CTP (Perkin Elmer).

### ***H3/H4 tetramer pull-down assay***

3 $\mu$ g of recombinant Xenopus H3/H4 tetramer were incubated with 1.25  $\mu$ M of recombinant Flag Set2 1-261, 1-476 TEV and 1-618TEV proteins in 20 $\mu$ l 1xHMT reaction system containing 50 mM Tris.HCl

pH 8.0, 300 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM DTT and 5 % Glycerol for 30 minutes at 30°C. After 30 minutes of pre-incubation, the reaction was expanded up to 120µl by adding 98.4µl of 1xHMT buffer, 0.2µl of 10% NP40, 1.4µl of 5M NaCl and 10µl pre-equilibrated 50/50 slurry of anti-Flag M2 affinity gel beads (Sigma) for 2 more hours' incubation at 4°C. The final pull-down reaction system now contained 1xHMT buffer supplemented with 150mM NaCl and 0.01% NP40. After 2 hours of incubation, the beads were washed for 3 times by 500µl of 0.15M GST buffer containing 150mM NaCl 20mM Tris-HCl pH7.5, 0.01% NP40 and 10mM beta-ME then eluted by 10µl of 3xSDS buffer for 15minutes at RT. The eluted proteins were boiled for 3 minutes and resolved by 10% SDS-PAGE. Western blotting was performed to detect pull-downed proteins by using anti-Flag HRP, anti-H3, and anti-H4 antibody.

Table 2.1 Plasmid List

| <b>Plasmid</b>  | <b>Backbone</b> | <b>Description</b>   | <b>Source</b> |
|-----------------|-----------------|--|---------------|
| <b>pBL196</b>   |                 | <i>pRET-N-GST</i>  | Li Lab        |
| <b>pBL532</b>   | pBP-HFT         | <i>pBacPAK-8-N-HisFlag-TEVsite</i>   | Li Lab        |
| <b>pBL645</b>   | pBS-105         | <i>pBS-216L-16X</i>  | Li Lab        |
| <b>pBL647</b>   | pBS-305         | <i>pBS-153C-16X</i>  | Li Lab        |
| <b>pBL648</b>   | pBS-405         | <i>pBS-147-16X</i>   | Li Lab        |
| <b>pBL766</b>   | pBL765          | <i>pBSKO-NgoM4-Bts1-2-3</i>  | Li Lab        |
| <b>pBL773</b>   | pBL766          | <i>pBSaK-Rpb1-KSn-Linker</i>   | Li Lab        |
| <b>pBL855</b>   |                 | <i>pYIA</i>  | John Stiller  |
| <b>pBL870</b>   | pCR-Blunt       | <i>pCR-B-CTD-fusion-Up</i>   | Li Lab        |
| <b>pBL871</b>   | pCR-Blunt       | <i>pCR-B-CTD-fusion-Down</i>   | Li Lab        |
| <b>pBL872</b>   | pCR-Blunt       | <i>pCR-B-yeast Set2-SET (Xho1) for CTD fusion</i>                          | this study    |
| <b>pBL875</b>   | pBL733          | <i>pBSaK-KSn-CTD fusion (Xho1) HA</i>                                      | this study    |
| <b>pBL876</b>   | pBL875          | <i>pBSaK-KSn-CTD fusion Set2-SET(Xho1)-HA</i>                              | this study    |
| <b>pWY001</b>   | pBL196          | <i>pRET-GST-Set2 1-261</i>   | this study    |
| <b>pWY001/A</b> | pBL196          | <i>pRET-GST-Set2</i>   | this study    |
| <b>pWY005</b>   | pBL532          | <i>pBP-HFT-Set2</i>  | this study    |
| <b>pWY037</b>   | pBL532          | <i>pBP-HFT-Set2 1-261</i>  | this study    |
| <b>pWY038</b>   | pBL532          | <i>pBP-HFT-Set2 1-618</i>  | this study    |
| <b>pWY039</b>   | pCR-Blunt       | <i>pCR-B-Set2 FL (Xho1) for CTD fusion</i>                                 | this study    |
| <b>pWY040</b>   | pCR-Blunt       | <i>pCR-B-Set2 1-618 (Xho1) for CTD fusion</i>                              | this study    |
| <b>pWY041</b>   | pBL875          | <i>pBSaK-KSn-CTD-fusion-Set2 FL(Xho1)-HA</i>                               | this study    |
| <b>pWY042</b>   | pBL875          | <i>pBSaK-KSn-CTD-fusion-Set2 1-618(Xho1)-HA</i>                            | this study    |
| <b>pWY043</b>   | pBL855          | <i>pYIA CTD-fusion-Set2 (KpnI/SnaBI)</i>                                   | this study    |
| <b>pWY044</b>   | pBL855          | <i>pYIA CTD-fusion-Set2 1-618 (KpnI/SnaBI)</i>                             | this study    |
| <b>pWY045</b>   | pBL855          | <i>pYIA CTD-fusion-Set2 1-261 (KpnI/SnaBI)</i>                             | this study    |
| <b>pWY046</b>   | pBL855          | <i>pYIA CTD-fusion(KpnI/SnaBI)</i>   | this study    |
| <b>pWY050</b>   | pBL532          | <i>pBP-HFT-Set2 1-618 TEV</i>  | this study    |
| <b>pWY051</b>   | pBL532          | <i>pBP-HFT-Set2 1-476 TEV</i>  | this study    |
| <b>pWY061</b>   | pBL532          | <i>pBP-HFT-Set2 SET-G5-SRI TEV (<math>\Delta</math>262-618)</i>            | this study    |
| <b>pWY062</b>   | pBL532          | <i>pBP-HFT-Set2 1-302 TEV</i>  | this study    |
| <b>pWY063</b>   | pBL532          | <i>pBP-HFT-Set2 1-338 TEV</i>  | this study    |
| <b>pWY064</b>   | pBL532          | <i>pBP-HFT-Set2 1-386 TEV</i>  | this study    |
| <b>pWY065</b>   | pBL532          | <i>pBP-HFT-Set2 1-415TEV</i>   | this study    |
| <b>pWY066</b>   | pBL532          | <i>pBP-HFT-Set2 1-733 TEV (FL)</i>   | this study    |
| <b>pWY069</b>   | pBL354          | <i>pRS415-pGAL1-HA-Set2</i>  | this study    |
| <b>pWY080</b>   | pRS416          | <i>pRS416-ADH1pro-Set2-CYC1ter</i>   | Brian Strahl  |
| <b>pWY084</b>   | pWY080          | <i>pRS416-ADH1pro-ADPT2-FLAG-CYC1ter</i>                                   | this study    |
| <b>pWY087</b>   | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 SET-G5-SRI-TEV(<math>\Delta</math>262-618)</i> | this study    |
| <b>pWY088</b>   | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 FL</i>   | this study    |

|               |        |  |            |
|---------------|--------|--|------------|
| <b>pWY091</b> | pRS415 | <i>pRS415-ADH1pro-ADPT2-FLAG-Set2FL-CYC1ter</i>    | this study |
| <b>pWY101</b> | pBL532 | <i>pBP-HFT-Set2 FL(BglII/BamHI) (B/B)</i>          | this study |
| <b>pWY102</b> | pBL532 | <i>pBP-HFT-Set2 ΔA-B (Δ262-338)</i>                | this study |
| <b>pWY103</b> | pBL532 | <i>pBP-HFT-Set2 ΔA-C (Δ262-386)</i>                | this study |
| <b>pWY104</b> | pBL532 | <i>pBP-HFT-Set2 ΔA-E (Δ262-476)</i>                | this study |
| <b>pWY105</b> | pBL532 | <i>pBP-HFT-Set2 ΔB (Δ303-338)</i>                  | this study |
| <b>pWY106</b> | pBL532 | <i>PBP-HFT-Set2 ΔC (Δ339-386)</i>                  | this study |
| <b>pWY107</b> | pBL532 | <i>pBP-HFT-Set2 ΔE (Δ416-476)</i>                  | this study |
| <b>pWY120</b> | pBL532 | <i>pBP-HFT-Set2 ΔA (Δ262-302)</i>                  | this study |
| <b>pWY121</b> | pBL532 | <i>pBP-HFT-Set2 ΔA-D (Δ262-415)</i>                | this study |
| <b>pWY122</b> | pBL532 | <i>pBP-HFT-Set2 ΔD (Δ387-415)</i>                  | this study |
| <b>pWY123</b> | pBL196 | <i>pRET-GST-Set2-SET-G5-SRI</i>                    | this study |
| <b>pWY126</b> | pWY074 | <i>pRS415-Set2 Pro/Ter-HA-Set2 ΔA (Δ206-303)</i>   | this study |
| <b>pWY127</b> | pWY074 | <i>pRS415-Set2 Pro/Ter-HA-Set2 ΔA-B (Δ206-338)</i> | this study |
| <b>pWY128</b> | pBL869 | <i>pRET-GST-Set2-SRI</i>                           | this study |
| <b>pWY129</b> | pWY084 | <i>pRS416-ADH1pro-FLAG-Set2 ΔA (Δ262-302)</i>      | this study |
| <b>pWY130</b> | pWY084 | <i>pRS416-ADH1pro-FLAG-Set2 ΔA-B (Δ262-338)</i>    | this study |
| <b>pWY131</b> | pBL944 | <i>pRS415-Set2 Pro/Ter-HA-Set2</i>                 | this study |
| <b>pWY136</b> | pWY084 | <i>pRS416-ADH1pro-FLAG-Set2 ΔC (Δ339-386)</i>      | this study |
| <b>pWY137</b> | pWY084 | <i>pRS416-ADH1pro-FLAG-Set2 ΔB (Δ303-338)</i>      | this study |
| <b>pWY138</b> | pWY084 | <i>pRS416-ADH1pro-FLAG-Set2 ΔD (Δ387-415)</i>      | this study |
| <b>pWY139</b> | pBL869 | <i>pRET-GST-Set2-477C</i>                          | this study |
| <b>pWY141</b> | pBL869 | <i>pRET-GST-Set2-262C</i>                          | this study |
| <b>pWY151</b> | pWY084 | <i>pRS416-ADH1pro-FLAG-Set2 ΔE (Δ416-476)</i>      | this study |
| <b>pWY152</b> | pWY084 | <i>pRS416-ADH1pro-FLAG-Set2 ΔA-E (Δ216-476)</i>    | this study |
| <b>pWY182</b> | pWY138 | <i>pRS416-ADH1pro-FLAG-Set2 ΔD-C201A</i>           | this study |
| <b>pWY184</b> | pWY137 | <i>pRS416-ADH1pro-FLAG-Set2 ΔB-C201A</i>           | this study |
| <b>pWY215</b> | pWY074 | <i>pRS415-Set2 Pro/Ter-HA-Set2 ΔD (Δ387-415)</i>   | this study |
| <b>pWY216</b> | pWY074 | <i>pRS415-Set2 Pro/Ter-HA-Set2 ΔB (Δ303-338)</i>   | this study |
| <b>pWY217</b> | pWY074 | <i>pRS415-Set2 Pro/Ter-HA-Set2 ΔC (Δ339-386)</i>   | this study |
| <b>pWY218</b> | pWY074 | <i>pRS415-Set2 Pro/Ter-HA-Set2 ΔE (Δ416-476)</i>   | this study |

Table 2.2 Strain List

| NAME     | Parental strain | Genotype   | Source       |
|----------|-----------------|--|--------------|
| YBL360   | BY4741          | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Rpb9-TAP::HIS3</i>  | Li Lab       |
| YCR377   | YCR376          | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i>  | Li Lab       |
| YYW010   | BY4742          | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN</i>  | Li Lab       |
| YYW026/A | BY4741          | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 HIS3::pGAL1-3HA-Set2</i>  | this study   |
| YYW037   | PH499           | <i>MATa ura3-52 lys2-801 ade2-101 his3-Δ200 leu2-Δ1Set2-C-TAP::TRP1</i>  | Li Lab       |
| YYW040   | BY4742          | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 Set2 1-618-C-TAP::URA3</i>  | this study   |
| YYW042   | BY4742          | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 HIS3::pGAL1-3HA-Set2 1-260-C-TAP::URA3</i>  | this study   |
| YYW044   | BY4742          | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 Set2 1-260-C-TAP ::URA3</i>   | this study   |
| YYW045   | YBL699          | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ΔRPH1::KAN ΔJHD1::Cre-LEU2 Set2 1-260-C-TAP::URA3</i>   | this study   |
| YYW087   | BY4742          | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 Set2-C-TAP::URA3</i>  | this study   |
| YYW114   | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY046(pYIA CTD-fusionKpnI/SnaBI Leu2)</i>   | this study   |
| YYW115   | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY045(pYIA CTD-fusion-Set2 1-261KpnI/SnaBI Leu2)</i>  | this study   |
| YYW117   | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY046(pYIA CTD-fusion KpnI/SnaBI Leu2)</i>  | this study   |
| YYW118   | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY045(pYIA CTD-fusion-Set2 1-261KpnI/SnaBI Leu2)</i>  | this study   |
| YYW120   | Z26             | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pRP112 (RPB1 CEN URA3)</i>   | John Stiller |
| YYW121   | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3ΔSET2::HPH + pRP112 (RPB1 CEN URA3 )</i>   | this study   |
| YYW122   | CKY283          | <i>MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128 ∂ gal10Δ56 rpb1Δ::CLONATMX Rpb3-TAP::KlacTRP1+pRP112 (RPB1 CEN URA3)</i>                              | C. Kaplan    |
| YYW123   | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY044(pYIA CTD-fusion-Set2 1-618 KpnI/SnaBI Leu2)</i>   | this study   |
| YYW124   | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY043(pYIA CTD-fusion-Set2 KpnI/SnaBI Leu2)</i>   | this study   |
| YYW125   | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY044(pYIA CTD-fusion-Set2 1-618 KpnI/SnaBI Leu2)</i>   | this study   |
| YYW126   | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY043(pYIA CTD-fusion-Set2 KpnI/SnaBI Leu2)</i>   | this study   |
| YYW127   | BY4742          | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 Set2 1-261-C-TAP ::URA3</i>   | this study   |
| YYW129   | YBL699          | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ΔRPH1::KAN ΔJHD1::Cre-LEU2 Set2 1-261-C-TAP::URA3</i>   | this study   |
| YYW134   | YYW122          | <i>MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128 ∂ gal10Δ56 rpb1Δ::CLONATMX Rpb3-TAP::KlacTRP1+pWY045 (pYIA CTD-fusion-Set2 1-261 KpnI/SnaBI Leu2)</i> | this study   |
| YYW136   | YYW122          | <i>MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128 ∂ gal10Δ56 rpb1Δ::CLONATMX Rpb3-TAP::KlacTRP1+pWY045 (pYIA CTD-fusion-Set2 1-618 KpnI/SnaBI Leu2)</i> | this study   |

|        |        |  |                |
|--------|--------|--|----------------|
| YYW138 | YYW122 | <i>MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128 ∂ gal10Δ56 rpb1Δ::CLONATMX Rpb3-TAP::KlacTRP1+pWY045 (pYIA CTD-fusion-Set2 KpnI/SnaBI Leu2)</i> | this study     |
| YYW162 | W303   | <i>MATa ade2 can1 his3 leu2 lys2 met15 trp1 ura3 spt16-11</i>  | D.<br>Stillman |
| YYW163 | W303   | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN</i>   | D.<br>Stillman |
| YYW165 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN ΔRCO1::cre-HIS(c.kl)</i>  | this study     |
| YYW800 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY129 (pRS416-ADH1pro-FLAG-Set2 ΔA(Δ262-302))</i>  | this study     |
| YYW801 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY137 (pRS416-ADH1pro-FLAG-Set2 ΔB(Δ303-338))</i>  | this study     |
| YYW802 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY136 (pRS416-ADH1pro-FLAG-Set2 ΔC(Δ339-386))</i>  | this study     |
| YYW803 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY138 (pRS416-ADH1pro-FLAG-Set2 ΔD(Δ387-415))</i>  | this study     |
| YYW804 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY151 (pRS416-ADH1pro-FLAG-Set2 ΔE(Δ416-476))</i>  | this study     |
| YYW805 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY152 (pRS416-ADH1pro-FLAG-Set2 ΔA-E(Δ216-476))</i>  | this study     |
| YYW806 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY087 (pRS416-ADH1pro-FLAG-Set2 SET-G5-SRI-TEV(Δ262-618))</i>  | this study     |
| YYW807 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY088 (pRS416-ADH1pro-FLAG-Set2 FL)</i>  | this study     |
| YYW808 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ΔSET2::KAN+pWY084 (pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>  | this study     |
| YYW809 | YYW162 | <i>MATa ade2 can1 his3 leu2 lys2 met15 trp1 ura3 spt16-11+pWY084 (pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>   | this study     |
| YYW810 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY084 (pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>  | this study     |
| YYW811 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY088 (pRS416-ADH1pro-FLAG-Set2 FL)</i>  | this study     |
| YYW812 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY129 (pRS416-ADH1pro-FLAG-Set2 ΔA(Δ262-302))</i>  | this study     |
| YYW813 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY137 (pRS416-ADH1pro-FLAG-Set2 ΔB(Δ303-338))</i>  | this study     |
| YYW814 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY136 (pRS416-ADH1pro-FLAG-Set2 ΔC(Δ339-386))</i>  | this study     |
| YYW815 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY138 (pRS416-ADH1pro-FLAG-Set2 ΔD(Δ387-415))</i>  | this study     |
| YYW816 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY151 (pRS416-ADH1pro-FLAG-Set2 ΔE(Δ416-476))</i>  | this study     |
| YYW817 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN +pWY182(pRS416-ADH1pro-FLAG-Set2 ΔD-C201A)</i>  | this study     |
| YYW818 | YYW163 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN +pWY184(pRS416-ADH1pro-FLAG-Set2 ΔB-C201A)</i>  | this study     |
| YYW819 | YYW165 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔRCO1::cre-HIS(c.kl) ΔSET2::KAN+pWY129(pRS416-ADH1pro-FLAG-Set2 ΔA(Δ262-302))</i>                                    | this study     |
| YYW820 | YYW165 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔRCO1::cre-HIS(c.kl) ΔSET2::KAN+pWY137(pRS416-ADH1pro-FLAG-Set2 ΔB(Δ303-338))</i>                                    | this study     |
| YYW821 | YYW165 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔRCO1::cre-HIS(c.kl) ΔSET2::KAN+pWY136(pRS416-ADH1pro-FLAG-Set2 ΔC(Δ339-386))</i>                                    | this study     |

|        |        |   |            |
|--------|--------|---|------------|
| YYW822 | YYW165 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔRCO1::cre-HIS(c.kl)</i><br><i>ΔSET2::KAN+pWY138(pRS416-ADH1pro-FLAG-Set2 ΔD(Δ387-415))</i>   | this study |
| YYW823 | YYW165 | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔRCO1::cre-HIS(c.kl)</i><br><i>ΔSET2::KAN+pWY151(pRS416-ADH1pro-FLAG-Set2 ΔE(Δ416-476))</i>   | this study |
| YYW824 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY131(pRS415-Set2 Pro/Ter-HA-Set2)</i>                                    | this study |
| YYW825 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY126(pRS415-Set2 Pro/Ter-HA-Set2 ΔA( Δ206-303))</i>                      | this study |
| YYW826 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY127(pRS415-Set2 Pro/Ter-HA-Set2 ΔA-B( Δ206-338))</i>                    | this study |
| YYW827 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY215(pRS415-Set2 Pro/Ter-HA-Set2 ΔD( Δ387-415))</i>                      | this study |
| YYW828 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY216(pRS415-Set2 Pro/Ter-HA-Set2 ΔB(Δ303-338))</i>                       | this study |
| YYW829 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY217(pRS415-Set2 Pro/Ter-HA-Set2 ΔC(Δ339-386) )</i>                      | this study |
| YYW830 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY218(pRS415-Set2 Pro/Ter-HA-Set2 ΔE(Δ416-476) )</i>                      | this study |
| YYW831 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY069(pRS415-pGALI-HA-Set2 )</i>  | this study |
| YYW832 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY052(pRS415 )</i>  | this study |
| YYW833 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i><br><i>ΔSET2::KAN+pWY091(pRS415-ADH1pro-ADPT2-FLAG-Set2</i><br><i>FL-CYCIter )</i>           | this study |
| YYW834 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY084(pRS416-ADH1pro-ADPT2-FLAG-CYCIter)</i>                 | this study |
| YYW835 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY088(pRS416-ADH1pro-FLAG-Set2 FL)</i>                       | this study |
| YYW836 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY129(pRS416-ADH1pro-FLAG-Set2 ΔA(Δ262-302))</i>             | this study |
| YYW837 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY137(pRS416-ADH1pro-FLAG-Set2 ΔB(Δ303-338))</i>             | this study |
| YYW838 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY136(pRS416-ADH1pro-FLAG-Set2 ΔC(Δ339-386))</i>             | this study |
| YYW839 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY138(pRS416-ADH1pro-FLAG-Set2 ΔD(Δ387-415))</i>             | this study |
| YYW840 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY151(pRS416-ADH1pro-FLAG-Set2 ΔE(Δ416-476))</i>             | this study |
| YYW841 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH</i><br><i>+pWY087(pRS416-ADH1pro-FLAG-Set2 SET-G5-SRI-TEV(Δ262-618))</i> | this study |
| YYW842 | YYW132 | <i>MATahis3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ΔRCO1::KAN</i><br><i>ΔSET2::URA3+pWY131(pRS415-Set2 Pro/Ter-HA-Set2)</i>                            | this study |
| YYW843 | YYW132 | <i>MATahis3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ΔRCO1::KAN</i><br><i>ΔSET2::URA3+pWY126(pRS415-Set2 Pro/Ter-HA-Set2 ΔA( Δ206-303))</i>              | this study |
| YYW844 | YYW132 | <i>MATahis3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ΔRCO1::KAN</i><br><i>ΔSET2::URA3+pWY127(pRS415-Set2 Pro/Ter-HA-Set2 ΔA-B( Δ206-338))</i>            | this study |



## **Chapter 3 PRECISE CONTACT OF SET2 WITH POL II CTD REPEATS IS ESSENTIAL FOR ITS HISTONE METHYLTRANSFERASE ACTIVITY**

### **Abstract**

The phosphorylated RNA polymerase II C-terminal domain (CTD) serves as a platform to recruit various factors including chromatin modifiers. In particular, the H3K36 methyltransferase Set2 has been shown to associate with the phosphorylated CTD through the SRI domain. Although it was shown that interaction between SRI and CTD is important for tri-methylated H3K36, the detailed underlying mechanism was unclear. In this study, we attempted to investigate if the phosphorylated Pol II CTD regulates Set2 activity beyond its recruitment. Surprisingly, we found the excessive amount of phosphorylated serine residues on CTD inhibited Set2 activity *in vitro*. The biophysical examination revealed that the additional phosphorylated CTD repeats collaterally occupied the surface of SRI where SRI contacts DNA. Finally, we determined that the minimal recognition unit of Set2 on fully phosphorylated CTD tail was three heptad repeats, and demonstrated that this minimal unit was sufficient for the Set2 recruitment without disrupting its catalytic activity. Since Pol II CTD utilizes repeating sequence as a scaffold for multiple factors, our results implicate that an organized spatial arrangement of these factors along CTD is necessary for accommodating their individual functions.

## Introduction

### RNA Pol II CTD phosphorylation cycle

RNA polymerase II (Pol II) is responsible to catalyze the transcription of protein encoding genes. C-terminal domain (CTD) of Pol II's largest subunit Rpb1 has been shown to be phosphorylated on Ser2, 5, and 7 residue of its highly conserved hepta-peptide  $Y_1S_2P_3T_4S_5P_6S_7$  during transcription (Bartkowiak and Greenleaf, 2011). In any transcription cycle, dynamic phosphorylation patterns of serine residues in CTD are essential to coordinate the recruitment of factors to ensure proper initiation, elongation and termination of mRNA. (Maniatis and Reed, 2002; Nechaev and Adelman, 2011) . RNA Pol II is recruited to the preinitiation complex (PIC) when CTD is hypophosphorylated (Lu et al., 1991). Following its binding to the promoter, Ser5 of yeast CTD is rapidly phosphorylated by either Kin28 (Cdk7 in human) (Feaver et al., 1991) or Cdk8 (Srb10 and Srb11 in human) (Liao et al., 1995). These two kinase complexes are thought to have partially overlapping roles in transcription (Liu et al., 2004) to facilitate the dissociation of RNA Pol II from the PIC and elicit promoter escape (Sogaard and Svejstrup, 2007). During transcription elongation, CTDK-1 and Bur1/Bur2 complexes are considered equally responsible for Ser2 phosphorylation during Pol II elongation stage. However, Bur1/Bur2 complex is thought to contribute Ser2P at early elongation stage, while deletion of Ctk1 was shown to reduce Serine phosphorylation in the downstream region by 90% (Qiu et al., 2009). At termination stage of transcription, CTD is de-phosphorylated by phosphatases recycled for the next round of transcription (Cho et al., 2001). Both Fcp1 and Ssu72 have been characterized as CTD phosphatase: Fcp1 is an essential for cell survival in yeast and it converts Pol II from phosphorylated IIO to unphosphorylated IIA (Archambault et al., 1997; Chambers and Dahmus, 1994) . Ssu72 has been shown to preferentially remove Ser5P from the CTD (Hausmann et al., 2005) and its activity is enhanced by the peptidyl prolyl isomerases Ess1 (Krishnamurthy et al., 2009).

### The CTD code

Definition of the CTD code: One of the central questions regarding CTD regulation is how this simple sequence repetition orchestrates so many phosphorylated CTD association factors (PCAPs) with broadly diversified functions in a precise temporal and special manner. A “CTD code” model has been proposed for quite a long time that different combinations of CTD modification guide the recruitment of specific

binding factor at the appropriate time rather than carrying all of them throughout the transcription process (Buratowski, 2003). In this chapter we mainly discuss CTD Serine phosphorylation.

Reading CTD code: Phosphorylated CTD binding factors have been shown to recognize specific CTD phosphorylation signals (on Ser2, Ser5, Ser7, Tyr1 or Thr4). For instance, Set1-mediated H3K4 methylation mostly enriches at promoter regions to facilitate Pre-initiation Complex (PIC) assembly and gene activation (Lauberth et al., 2013). This is due to the recognition of Ser5P CTD which is unique at 5' end of gene (Ng et al., 2003). Other examples including integrator subunit, Int11, which recognizes a distinguished Ser7P/Ser2P double mark on CTD (Egloff et al., 2010) and RNA-binding protein Nrd1 which interacts with Ser5P CTD and requires a Ser5P-Pro6 bound in *cis* conformation (Kubicek et al., 2012). Alternatively, cooperative binding may also contribute to p-CTD recruitment. For instance, CTD bound transcription termination factors Rtt103 and Pcf11 prefer longer (four instead of two) heptad peptide repeats because two of each factor cooperatively bind to neighboring CTD repeats respectively (Lunde et al., 2010). DNA damage response requires CTD bound Set2. Interestingly, its enzymatic activity is dispensable for this function pathway. In this case, Set2 may function to coordinate CTD association of a subset of DNA damage response factors (Winsor et al., 2013).

### **Set2 is one of Phosphor-CTD-associating proteins (PCAPs)**

To date, a large number of PCAPs have been found to associate with phosphorylated CTD tail including elongation factors (Corden and Patturajan, 1997), mRNA capping enzymes (Cho et al., 1997; McCracken et al., 1997a), splicing enzymes (de la Mata and Kornblihtt, 2006), editing factors (Ryman et al., 2007), polyadenylation and 3' cleavage factors (McCracken et al., 1997b), and transcription termination factors (Gudipati et al., 2008; Vasiljeva et al., 2008) (see review (Eick and Geyer, 2013)). Yeast H3K36 methyltransferase Set2 (Strahl et al., 2002), has been shown to associate with elongating RNA Pol II through the interaction between SRI domain and p-CTD (Kizer et al., 2005; Li et al., 2003a). Similarly, Set2 orthologs in higher eukaryote, human Set2/Hypb and fly Hypb, also require association with phosphorylated Pol II through conserved SRI domain to achieve transcription-coupled H3K36me3 (Bell et al., 2007; Edmunds et al., 2008; Sun et al., 2005).

### **The structural insights into Set2/p-CTD interaction**

Set2 and its higher eukaryotes orthologs Setd2/Hypb utilize conserved Set2-Rpb1 interacting domain

to associate with phosphorylated CTD (Kizer et al., 2005; Sun et al., 2005). NMR studies show that both Set2 and Setd2/Hypb SRI domain form three-helix left-handed bundles, which is different from other CTD binding domains (Li et al., 2005; Meinhart et al., 2005; Vojnic et al., 2006). The proper CTD phosphorylation patterns for Set2 binding have also been studied. It was shown that Set2 SRI domain interacts with phosphorylated Ser2/Ser5 CTD peptide comprising two-heptad repeats and three flanking NH<sub>2</sub>-terminal residues (2,5,2,5 peptide). On the other hand, human SRI interacts best with phosphorylated Ser2,5 CTD peptides that contains at least two complete repeats (2,5,2,5,2,5 peptide, 2,5,2,5 peptide, and 5,2,5,2 peptide), with several folds weaken affinity towards 2,5,2 and 5,2,5 peptides (Li et al., 2005; Vojnic et al., 2006)

### **Set2 interaction with p-CTD is important for H3K36 methylation**

Set2/p-CTD binding is required for H3K36 tri-methylation (Youdell et al., 2008), possibly because unbound Set2 will be degraded rapidly (Fuchs et al., 2012). However multiple lines of evidence also suggest that SRI/p-CTD contact may play additional roles besides affecting Set2 protein stability. Over-expressing WT Set2 in  $\Delta CTK1$  cells which lacked CTD phosphorylation restores occupancy of Set2 but not H3K36me<sub>3</sub> on some genes (Youdell et al., 2008). This is consistent with what we observed: over-expressing WT Set2 was insufficient to tri-methylate H3K36 in  $\Delta PAF1$  and  $\Delta CDC73$  yeast strains though Set2 protein level is normal (Fig 2.2 B). Most intriguingly, in our Rpb1-Set2 fusion system, only the CTD-Set2-FL but not CTD-SET resumes H3K36me<sub>3</sub> although both were recruited to nucleosome by fusion to elongating Pol II (Fig 2.1).

H3K36me<sub>2/3</sub> subsequently recruit histone deacetylase complex (HDAC) Rpd3S (Li et al., 2007b) to suppress intragenic cryptic transcription at 3' of genes (Carrozza 2005). Loss of CTD phosphorylation correlates with cryptic transcription phenotype. For instance, deletion of Ctk1, the catalytic subunit (Ctk1) of CTD kinase CTDK-1 complex, or Bur2, the cyclin factor within Bur1 complex, causes cryptic transcription (Chu et al., 2007; Youdell et al., 2008). Deletion of one of Paf1 subunit Ctr9 causes cryptic transcription. (Li et al., 2009a; Youdell et al., 2008). Interestingly, Rpb3S complex can directly associate with Ser5 phosphorylated Pol II CTD. However, its histone deacetylase activity still requires H3K36 methylation (Govind et al., 2010).

Though *in vitro* CTD peptide binding studies help discover the CTD binding signals of PCAPs. Little is known as to whether CTD binding affects activities of PCAPs. We started our study by testing whether phosphorylated CTD (p-CTD) allosterically stimulates yeast Set2 methyltransferase activity on

H3K36. Multiple lines of evidence suggest SRI/p-CTD interaction is required for tri-methylated H3K36 in cells (Youdell et al., 2008) (result in Chapter 2). Surprisingly, we observed that phosphorylated FL CTD tail, instead of stimulating Set2 catalysis, inhibited Set2 activity by competing off the interaction of SRI with nucleosomal DNA. Another interesting finding was that Set2 only required three out of the 28 CTD heptad peptide repeats to achieve full enzymatic activity.

Our unexpected result suggests a new CTD binding model in which the number of phosphorylated CTD Serine residues should be controlled for the proper function of CTD binding factors. In particular for Set2, it may actively search for the specific CTD location with correct phosphorylated Serine number to balance CTD recruitment and enzymatic activity.

## Results

### Set2 middle domain regulates the binding of SRI to phosphorylated CTD

It has been shown that SRI domain interaction with RNA Pol II CTD is important for H3K36me (Kizer et al., 2005; Li et al., 2003a; Youdell et al., 2008). We wanted to test whether this interaction is regulated by other domain of Set2. We wanted to purify TAP-tagged Set2-FL and SRI deletion complex to compare their Pol II binding properties. Consistent with previous results, silver staining showed that the C-terminal TAP-tagged Set2-FL was co-purified with Pol II complex (Krogan et al., 2003; Li et al., 2003a) while the TAP- Set2  $\Delta$ SRI abolished this interaction (Fig 3.1 B). In addition, consistent with previous observations, TAP- Set2  $\Delta$ SRI lost both H3K36me<sub>2</sub> and H3K36me<sub>3</sub> (Fig 2.4 A).

In search for additional Set2 regions that regulate SRI/CTD interaction, we generated a series of recombinant GST-tagged Set2 N-terminal truncation and test their interaction with biotinylated CTD peptides composing of four-heptad repeats immobilized on streptavidin beads (Fig 3.1A) (Li et al., 2003a). Results of pull-down assay showed that WT Set2 specifically interacts with Ser2 or Ser5 phosphorylated CTD. Notably, SRI bound p-CTD with similar affinity and specificity to WT, suggesting this domain is sufficient for Set2/p-CTD interaction. Surprisingly, Set2 longer N-terminal truncations display reduced CTD binding specificities. N-terminal truncated mutants a.a. 262-733 and a.a. 476-733 contain full length or partial Set2 middle domain, respectively. This domain (AID) has been shown to auto-inhibit Set2 activity (Fig 2.5E). These mutants also prefer phosphorylated CTD peptides. However, they also slightly bind to un-phosphorylated CTD (Fig 3.1C). This could be explained by the fact that the middle domain contains a WW domain which recognizes Proline-rich ligand (Macias et al., 2002) and that un-phosphorylated CTD which is considered as proline-rich protein sequences (repetitive PxY core). In fact, unphosphorylated CTD has been shown to be associated by other WW domain containing proteins (Chang et al., 2000). Moreover, in SET-T-SRI fusion protein where the entire middle domain was deleted, we observed significantly reduced phosphorylated CTD binding (Fig3.1 C). Remarkably, SET domain directly inhibited SRI/p-CTD contact (compared GST-619-733 (SRI) with GST- $\Delta$ 262-618). Collectively, our CTD binding results show that the middle AID and SET domains controlled Set2 SRI binding to p-CTD, suggesting cross-talk of Set2 domains is essential for proper H3K36 methylation.

### Phosphorylated RNA Pol II complex inhibits Set2 activity

Multiple lines of evidence suggest that SRI/p-CTD contact may play additional roles in H3K36me

besides affecting Set2 protein recruitment and stability. First, Over-expressing WT Set2 in  $\Delta CTK1$  cells which lacked CTD phosphorylation restores occupancy of Set2 but not H3K36me3 on some genes (Youdell et al., 2008). In Chapter 2, we also observed that in our Rpb1-Set2 fusion system, only the CTD-Set2-FL but not CTD-SET resumes H3K36me3 although both were recruited to nucleosome by fusion to elongating Pol II (Fig 2.1). In addition, over-expressing WT Set2 was insufficient to tri-methylate H3K36 in  $\Delta PAF1$  and  $\Delta CDC73$  yeast strains though Set2 protein level is normal (Fig 2.2 B).

To study the detailed mechanism of how p-CTD regulates Set2 activity, we tried to test the whether phosphorylated Pol II affects Set2 activity *in vitro*. The native Pol II complex from yeast by TAP purification was purified via its Rpb9 subunit (Carey et al., 2006) and the complex as phosphorylated *in vitro* using FLAG tagged recombinant yeast Pol II kinase CTDK-1 complex (Sterner et al., 1995). Silver staining of phosphorylated Pol II showed the Rpb1 subunit migrated more slowly than the unphosphorylated sample, indicating successful phosphorylation (Figure 3.2 B). Remaining kinase complexes were removed by anti-FLAG beads. Surprisingly, phosphorylated Pol II inhibited Set2 activity in a dose-dependent manner while un-phosphorylated Pol II had no significant effect (Figure 3.3 C).

CTD fused Set2-FL/Pol II complex was also purified the through TAP-tagged Rpb3 from yeast (Chapter 2 Fig 2.1C). The purified fusion Pol II complexes were further phosphorylated by recombinant yeast CTDK-1 complex in presence of ATP (Figure 3.2 D). The activity of phosphorylated (with ATP) or unphosphorylated (without ATP) fusion CTD-Set2-FL/Pol II complex was compared in HMT assay (Figure 3.2F). Set2 activity in phosphorylated CTD-Set2-FL/Pol II complex decreased dramatically, consistent with what we observed in WT Pol II complex (Fig 3.2 C). In contrast, Set2 retained its activity in unphosphorylated CTD-Set2-FL/Pol II complex (Figure 3.2F). This difference is not likely due to phosphorylation of Set2 itself by CTDK-1 because the activity of recombinant Set2 remain the same regardless of its phosphorylation status (Figure 3.2F). Thus the above independent assays both indicate that, in contrary to current model, *in vitro* phosphorylated Pol II inhibits Set2 activity on nucleosome.

### **Ser2 Phosphorylated CTD peptides do not inhibit Set2 activity**

To further investigate how phosphorylated Pol II inhibits Set2 activity, we tested if four-heptad Ser2P CTD peptide had the same inhibitory effect. The CTD peptide contained similar number of

phosphorylation sites and had been identified as Set2-interacting unit (Vojnic et al., 2006). We also showed by pull-down assay that it can efficiently associate with Set2 (Fig 3.1 C). In contrast to phosphorylated Pol II, neither Ser2 phosphorylated nor un-phosphorylated CTD peptides inhibited Set2 activity. This result was unaffected even when the molar ratio between peptides and Set2 reached to 125::1 (Figure 3.3).

Considering CTDK-1 is promiscuous to phosphorylate CTD tail *in vitro* (Jones et al., 2004), the discrepancy of HMT results of Pol II and p-CTD peptide may be caused by different phosphorylation pattern of these two proteins. For example, phosphorylation of non-consensus region of CTD tail, amino acid residues other than Ser2 in heptad repeats (e.g. Tyr1, Thr4 or Ser7), or the overall amount of Ser2, 5 phosphorylation on CTD tails or peptide. Moreover, RNA Pol II contains 12 subunits (Bushnell and Kornberg, 2003). Although we only observed major migration shift of Rpb1 subunit after kinase reaction (Fig 3.2 B and E), we could not rule out the possibility that in our *in vitro* kinase system, unexpected promiscuity of CTDK-1 complex phosphorylated other subunits of Pol II which cause inhibition of Set2.

### **Phosphorylated GST-CTD inhibits Set2 activity through the SRI domain.**

To study the phosphorylation-dependent inhibitory of CTD in more detail, we constructed an engineered CTD tail expression system that recapitulated the repetitive sequences of WT PolIII CTD tail while enabled us to mutate any position along CTD. The vector contained 28-heptad repeats (compared with 26-heptad repeats in natural yeast CTD) and non-consensus sequences in natural CTD tail (Fig 3.4 A to D). The Ser2 or Ser5 within each seven-heptad block can be engineered to single or double alanine mutation (Figure 3.4 C and see Material & methods for details).

N-terminal GST-tagged FL WT CTD tail was purified from *E. coli* and subjected to phosphorylation using FLAG tagged Bur1 complex, a known CTD kinase (Fig 3.5 A). Coomassie blue staining showed that CTD migration shift up after Bur1 treatment (Figure 3.5 C), indicating successful phosphorylation. Bur1 complex was then removed by adding anti-FLAG beads to the system. We tested the inhibition of Set2 activity used the re-purified phosphorylated CTD tail (Figure 3.5 B). As shown in Figure 3.5D, the *in vitro* phosphorylated CTD tail inhibits Set2 activity in a dose-dependent manner, similar to what we observed before (Figure 3.2). In contrast, this inhibitory effect did not happen on Set2 a.a. 1-261 which was known to lack p-CTD interaction (Figure 3.5 D). Thus, the inhibition of phosphorylated Pol II on Set2 activity depends on SRI/ p-CTD tail interaction.

We also directly compared the inhibitory effect of GST- CTD tail and CTD peptides both with

four-heptad repeats on Set2. In a system coupling phosphorylation and HMT reaction, we found that phosphorylated CTD tail but not CTD peptides inhibited Set2 activity while unphosphorylated versions of neither cause inhibition (Fig 3.5 E). Notably, we ruled out the possibility that phosphorylation on GST tag may cause defect in H3K36me by showing that Set2 had comparable activity in presence of GST and CTD peptide (Figure 3.5 E).

### **Binding to phosphorylated GST-CTD neutralizes all SRI functions.**

We have shown that the middle domain (aa262-618) and SRI domain (aa619-733) antagonize each other in regulating Set2 activity (Fig 2.5 E). We want to further investigate whether the interplay between catalytic domain and auxiliary domains is also regulated by binding to CTD tail. To this end, we first tested how p-CTD tail influenced Set2 activity on histone H3 and observed p-CTD neutralizes SRI function when the middle domain is absent.

In this system biotin conjugated H3 peptides were immobilized by streptavidin coated magnetic beads (Figure 3.6 A). WT Set2, SET domain (a.a.1-261), middle domain deletion (SET-T-SRI) and a partial middle domain deletion  $\Delta B$  ( $\Delta$  a.a. 303-338) were incubated with the substrate in presence or absence of p-CTD tails (Figure 3.6 B). After reaction, the methylated H3 peptide was enriched on magnetic beads and incorporated  $^3\text{H}$  was measured. WT Set2 had constant low activity while SET had constant high activity on histone H3 (Figure 3.6 C). This is probably due to the fact that catalytic center of WT Set2 cannot access H3 substrate because of its C-terminal region (including AID and SRI domain). Notably, despite adding p-CTD, WT Set2 remained inactive on H3. This is consistent with the notion that other histone contact surface or DNA is required for Set2 activity on nucleosome (Fig 2.3E) and (Du et al., 2008). In contrast, p-CTD rescued the activities of the two middle domain truncations to SET levels. We hypothesize that free SRI may interfere with SET interaction with its substrate and impede its activity. SRI/p-CTD interaction may remove this blockage and rescue H3K36me (Figure 3.6 C).

We then tested how p-CTD tail influenced activity of WT Set2 and middle domain deletion (SET-T-SRI) on nucleosome context and found p-CTD does not fully repress the activity of middle domain deletion. Both Set2-FL and Set2-TEV-FL showed activity on Hela LON nucleosomes. Notably, SET-T-SRI is super-active, possibly due to lacking of repression from the middle domain (Fig 2.5 E). Adding p-CTD completely inhibited Set2 FL activity while the activity of SET-T-SRI was reduced but not totally inhibited (Fig 3.6 D).

To directly test whether the binding of Set2 to p-CTD interferes with Set2 contact DNA and nucleosome, we performed gel shift assays (EMSA) using  $^{32}\text{P}$  labeled double-strand DNA or reconstituted nucleosomes as probes. As shown in Figure 3.7 A, adding increasing amount of p-CTD into the pre-incubated Set2 and  $^{32}\text{P}$  labeled DNA disrupted the interaction between them and caused release of increasing amount of free DNA. p-CTD also interfered with Set2 binding to nucleosome. Free nucleosomes were released from upper-shifted Set2/Nucleosome species when p-CTD was presented (Figure 3.7 B Lane 5, 6, 7). In contrast, four-heptad Ser2P CTD peptide did not affect Set2/Nucleosome interaction (Figure 3.7 B Lane 9, 10). The EMSA data supported the hypothesis that additionally phosphorylated CTD compete with Set2 binding to substrate and impede with it activity.

Collectively, SET catalytic domain is regulated by both middle and SRI domains *in cis*. And the SRI/p-CTD tail interaction neutralized all SRI functions and left the SET domain fully repressed by the auto-inhibitory middle domain.

### **Extra copies of phosphorylated heptapeptide CTD collaterally occupies at the DNA binding surface of SRI.**

Set2 has been shown to prefer nucleosome (Strahl et al., 2002). We found a novel DNA binding ability of SRI (Fig 3.8B). Moreover, we showed that p-CTD tail neutralized all SRI functions and left SET domain fully suppressed (Fig 3.6). We hypothesize that loss of SRI/DNA binding is the direct cause of Set2 activity loss on nucleosome. Though structures of p-CTD peptide binding surfaces on both yeast and human SRI domain have been solved (Li et al., 2005; Vojnic et al., 2006), the p-CTD tail interacting residues of Set2 are still unclear.

We examined the detailed SRI domain conformational changes upon binding to phosphorylated CTD tail (p-CTD tail) using Deuterium exchange mass spectrometry (DXMS) technology which measures the hydrogen/deuterium exchange rates at each residue to reflect solvent accessibility at the region (Engen, 2009). By comparing the deuterium exchange rates (H/D) of each residue in free SRI to that bound to p-CTD tail, we obtained the heat map of the H/D rate changes that pinpointed amino acid residues that being either more protected or more exposed upon SRI/p-CTD interaction (Fig 3.8 A). Several highly protected regions upon SRI/p-CTD binding were identified and almost all of the strongly protected residues located in the SRI domain (data not shown). By mapping protected SRI residues from H/D exchange assay onto published yeast SRI structure model (Vojnic et al., 2006), we observed much larger p-CTD tail binding surfaces (Fig 3.8 C Left panel p-CTD FL/Set2, dark blue area according to the scale

bar in Fig 3.8A) than that was found in previous p-CTD peptide binding assay (Fig 3.8 C, right panel p-CTD peptide /Set2, purple or green area according to residue marks in Fig 3.8 A).

Multiple residues with potential DNA binding abilities were found in SRI domain (Fig 3.8A cyan *crosses* on top of amino acid sequence). Strikingly, when those predicted DNA binding residues were mapped onto published SRI/p-CTD peptide structure, we found: 1) None of predicted DNA binding residues overlaps with p-CTD peptide binding surface (Fig 3.8 C, compare surfaces in middle panel and colored surfaces in right panel). 2) Besides the intrinsic p-CTD peptide binding surface, p-CTD tail introduces additional contact surface into SRI (Notice that the blue surfaces included most of green and purple surfaces). 3) Most of the predicted DNA binding residues are shared by the additional binding surface introduced by p-CTD tail (Figure 3.8 C middle panel).

Based on above findings, we hypothesized that reason why only p-CTD tail but not p-CTD peptide can inhibit Set2 activity on nucleosome was because the additional phosphorylated Ser residues on p-CTD tail competed for binding surface of SRI with nucleosome which inhibited Set2 recruitment to nucleosomes.

### **Extra copies of phosphorylated heptapeptide CTD inhibits Set2 HMT activity.**

In our *in vitro* assay, Set2 activity was inhibited by heavily phosphorylated CTD. We hypothesized that optimal phosphorylation environment for each CTD bound Set2 may be required for the optimal enzyme activity. To test this, we utilized the engineered CTD system (Fig 3.4) to construct various length of heptad repeats and screened for optimal Set2 binding and activity. We first constructed the FL CTD tails that only containing 14 or 28 Ser2, 5 with the rest Ser residues being replaced by Ala (Fig 3.9 A). The purified CTD tails were phosphorylated (Data not shown) and added into HMT assay of recombinant Set2 (Figure 3.9 B). However, Set2 activity on nucleosome can not be rescued by these CTD with decrease phosphorylation (Fig 3.9 C) indicating seven-heptad WT repeat CTD which contains much more phosphorylated Serines than p-CTD peptide (14 compared to 4) is still sufficient to inhibit Set2 activity.

We then further reduced the total phosphorylation on CTD tail by mutated Ser2, 5 to Ala in individual heptad repeats. In this case, p-CTD with seven phosphorylated heptads still retains about 80% of its inhibitory ability, while p-CTD with four phosphorylated repeat started to show compromised inhibition on Set2 activity (Fig3.9 D). Further reduction to three, two or one phosphorylated heptads led to less than 20% of inhibition (Figure 3.9 D, E). And all S2, 5A mutant CTD (1\*(28)) did not

affect Set2 activity at all. Therefore, we concluded that p-CTD tail inhibition on Set2 was due to excessive amount of Ser2/5 phosphorylation exposed to the methyltransferase.

To rule out the possibility that loss of p-CTD inhibition was due to the loss of SRI/p-CTD interaction, we tested the interaction of Set2 with all p-CTD tail variants we used in HMT in a pull down assay (Fig 3.9F). We found that the minimum requirement for Set2 binding is three-heptad repeats and C1-1WT and C1-2WT CTD fail to interact with Set2 (Figure 3.9 G). This finding is also consistent with p-CTD peptide binding result that at least four phosphorylated Serine residues are required for Set2 binding. Collectively, we concluded the optimal p-CTD length for Set2 binding and activity is three phosphorylated heptad repeats. Higher phosphorylation density may cause inhibition of the enzyme.

## Discussion

In this study we used biochemical approaches to study how phosphorylated CTD code regulates Set2 methyltransferase activity. We found that, firstly, in contrast to factors like Rtt103 that prefer high phosphorylation density-prone CTD (Lunde et al., 2010), Set2 activity on nucleosome was inhibited by excess phosphorylation on CTD tail. In specific, Set2 required only three CTD heptad repeats for proper recruitment and function. Secondly, our study revealed an additional function of Set2 SRI domain. Besides recruiting the enzyme to p-CTD, SRI also fine-tuned the interaction among Set2, nucleosome DNA and p-CTD. We speculated that when excessive phosphorylated residues were presented, the interaction with p-CTD may compete with SRI bind to nucleosomal DNA thus compromised Set2 activity.

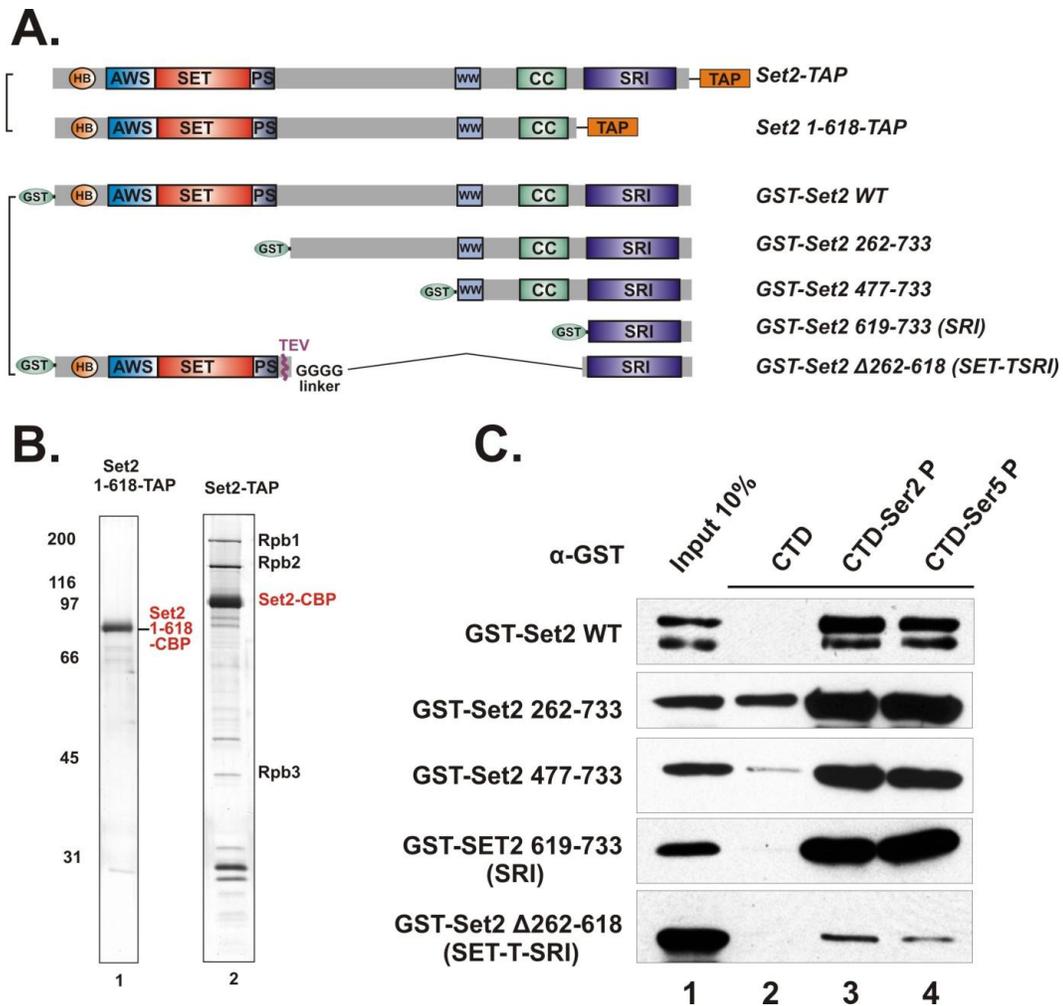
By comparing C1-3WT and C1-4WT CTD tails, we notice that the assuming the Serine phosphorylated is saturated after kinase reaction, the maximal number of phosphorylated Serine, 5 residues that do not affect Set2 activity is 7 because 8 phosphorylated Serine residues are able to cause about 50% decrease of Set2 activity. And the comparison between two-heptad WT repeats which does not bind to Set2 with previously used four-heptad Ser2P peptide suggests that the minimal binding module of CTD for Set2 is about four continuously phosphorylated Serine residues in di-heptad repeat flanked by the adjacent heptad residues.

Based on our *in vitro* results we argue that the improper ratio between phosphorylated CTD Serine residues and Set2 causes an unfavorable environment for Set2. However this observation may imply that the optimal phosphorylation density for each CTD bound Set2 is required for the H3K36 *in vivo*. It is

believed that in the physiology condition the phosphorylated CTD tail is co-occupied by various CTD binding factors to coordinate Pol II transcription and other functions. Meanwhile, multiple CTD phosphatases are also involved in balancing the phosphorylation density (Bataille et al., 2012; Phatnani and Greenleaf, 2006) presumably leaving very limited amount of the exposed phosphorylated serine residues. Yeast CTD contains 26 copies of heptad repeats giving 52 possible Ser2, 5 phosphorylation sites (Corden et al., 1995) while just 3-4 continuously phosphorylated Ser2 and 5 residues across two and half CTD heptad repeats are sufficient for SRI association (Li et al., 2003a; Vojnic et al., 2006). Thus, Set2 must locate either on an isolated CTD region or a CTD binding “module” formed by multiple neighboring binding factors where only proper phosphorylated Serine residue are exposed.

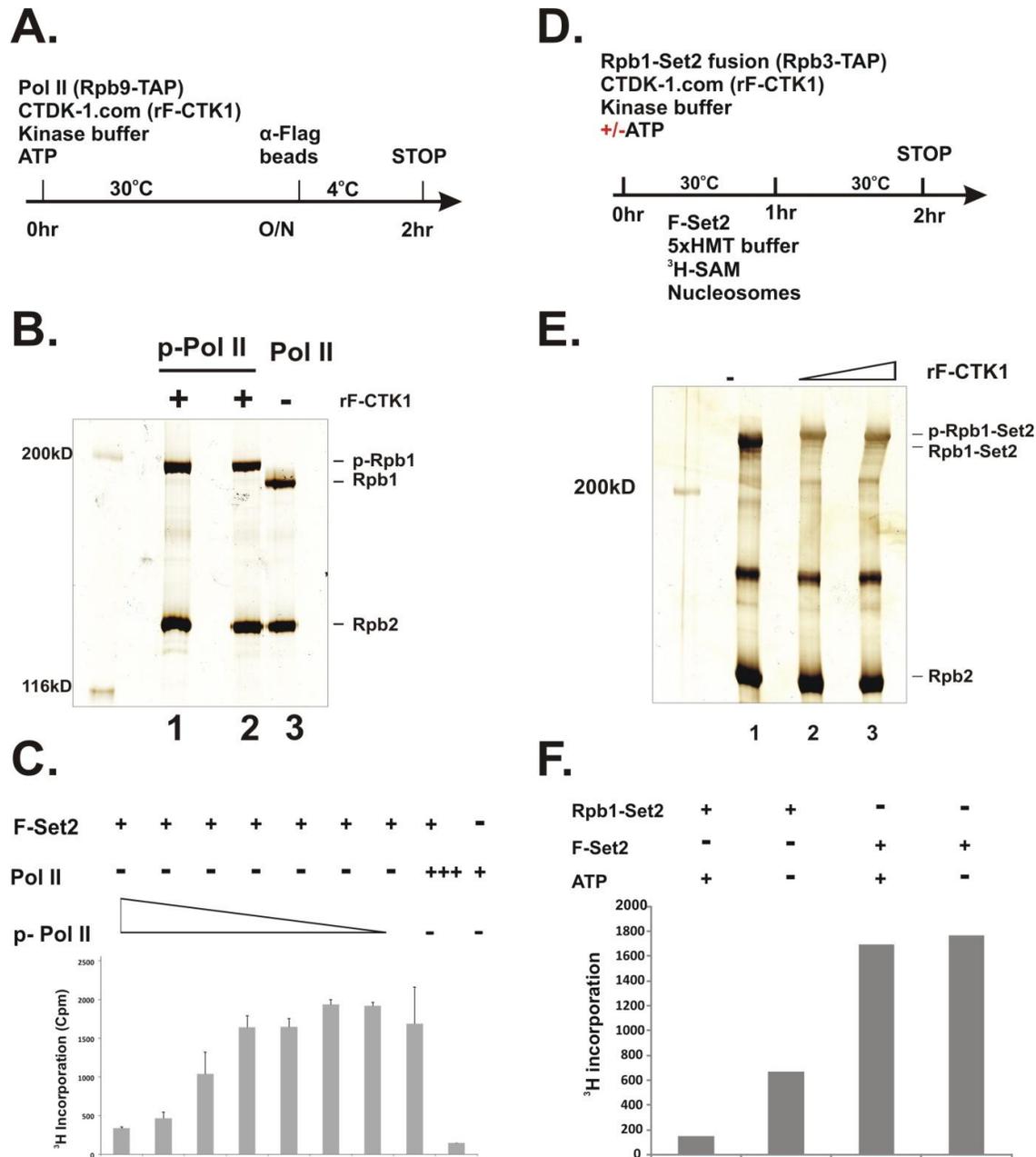
We speculate that there are two strategies for Set2 to find the correct CTD code balancing both CTD recruitment and enzyme activity. The random searching for the isolated p-CTD region with correct phosphorylation density is energetically inefficient and requires cooperation between Set2 and CTD phosphatase. Alternatively, Set2 can reside in a fixed CTD binding “module” formed by neighboring factors which constantly provides favored CTD phosphorylation environment.

Genetic studies show that overlong CTD tail has functional defect under stress, presumably because the additional distal heptad repeat competitively recruit the proximal repeat binding factors (Liu et al., 2010). Emerging evidence also suggests that a number of CTD associated factors follow a well-organized manner to cooperatively occupy the CTD for the DNA damage response and this does depend on Set2 binding (Winsor et al., 2013). Thus, the neighboring PCAPs and Set2 may form a specific subset of PCAP module to coordinate each other's function during transcription.



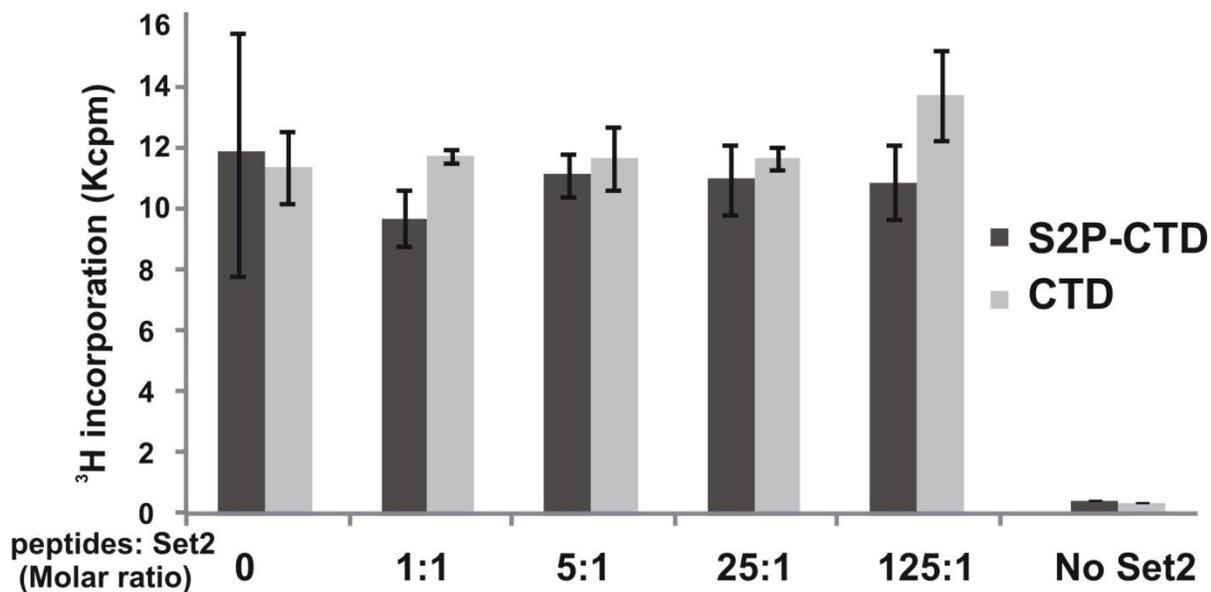
**Figure 3.1 Set2 middle domain regulates the binding of SRI to phosphorylated CTD.**

(A) The scheme of C-terminal TAP-tagged Set2 FL and SRI domain deletion (1-618-TAP). GST-tagged Set2 FL and mutants. (B) Silver staining showing TAP-purified FL Set2 but not Set2 1-618 associates with RNA Pol II complex. (C) *In vitro* pull-down assay using recombinant GST Set2 variants and biotinylated CTD peptides: unphosphorylated (CTD), phosphorylated on Ser5 (Ser5P) or Ser2 (Ser2P). Streptavidin-coated magnetic beads were used to pull down CTD peptide bound Set2 variants and Western blot analysis was performed using anti-GST antibody

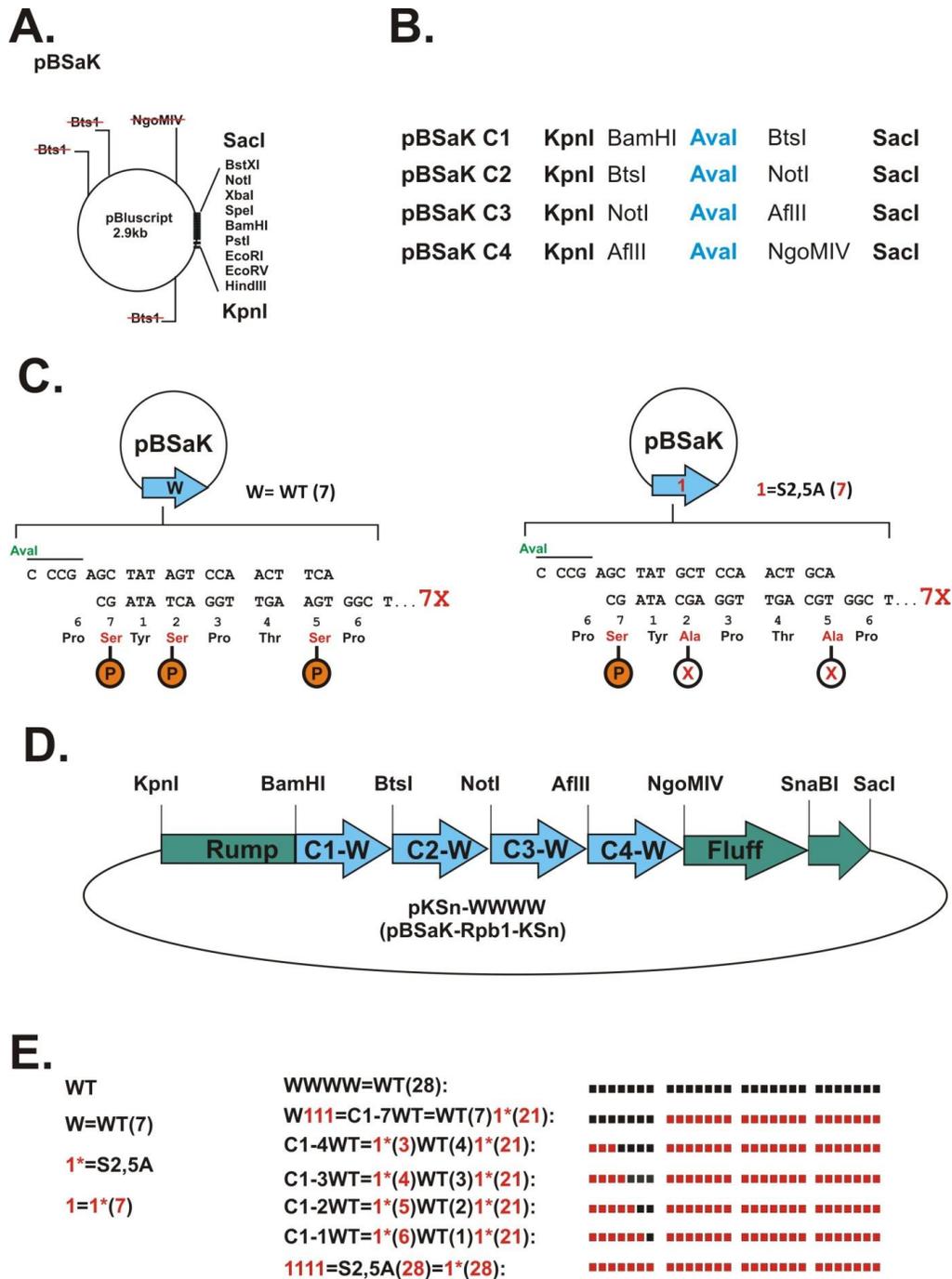


**Figure 3.2 Phosphorylated RNA Pol II complex inhibits Set2 activity.** (A) An experimental scheme of *in vitro* phosphorylation assay using baculovirus expressed FLAG-tagged CTD kinase CTDK-1 complex (rF-CTK1) to phosphorylate RNA Pol II from tandem purification through TAP-tagged Rpb9 subunit. Anti-FLAG beads were applied after phosphorylation reaction to remove rF-CTK1 from final products. (B) Silver staining showing the shifted migration of Rpb1 subunit from phosphorylation reaction (Lane 2) and anti-FLAG beads incubation (Lane 1). (C) HMT assay using HeLa LON nucleosomes as substrate to test the inhibitory effect of

phosphorylated RNA Pol II (p-Pol II) upon F-Set2. HMT activity was represented as mean  $\pm$  SEM, n=3. (D) An experimental scheme of testing the methyltransferase activity inhibition of Rpb1-fused Set2 by *in vitro* phosphorylation using baculovirus expressed FLAG-tagged CTDK-1 complex (rF-CTK1). The Rpb1-Set2 fusion RNA Pol II complex was purified through TAP-tagged Rpb3 subunit (Rpb3-TAP). Right after kinase reaction, Hela LON, 5xHMT buffer, and  $^3\text{H}$  labeled SAM were added to carry out methyltransferase reaction. The incorporated  $^3\text{H}$  was read by scintillation counter. (E) Silver staining of the shifted migration of phosphorylated Rpb1-Set2 (p-Rpb1-Ser2) by rF-CTK1. (Lane2 and 3 compared with Lane 1). (F) HMT assay illustrated in (A) showing the inhibited Rpb1-Set2 methyltransferase activity after kinase plus ATP treatment compared with no ATP treatment. The recombinant FLAG-tagged Set2 (F-Set2) incubated with kinase (with and without adding ATP) showing kinase reaction does not affect Set2 to inhibit methyltransferase activity.

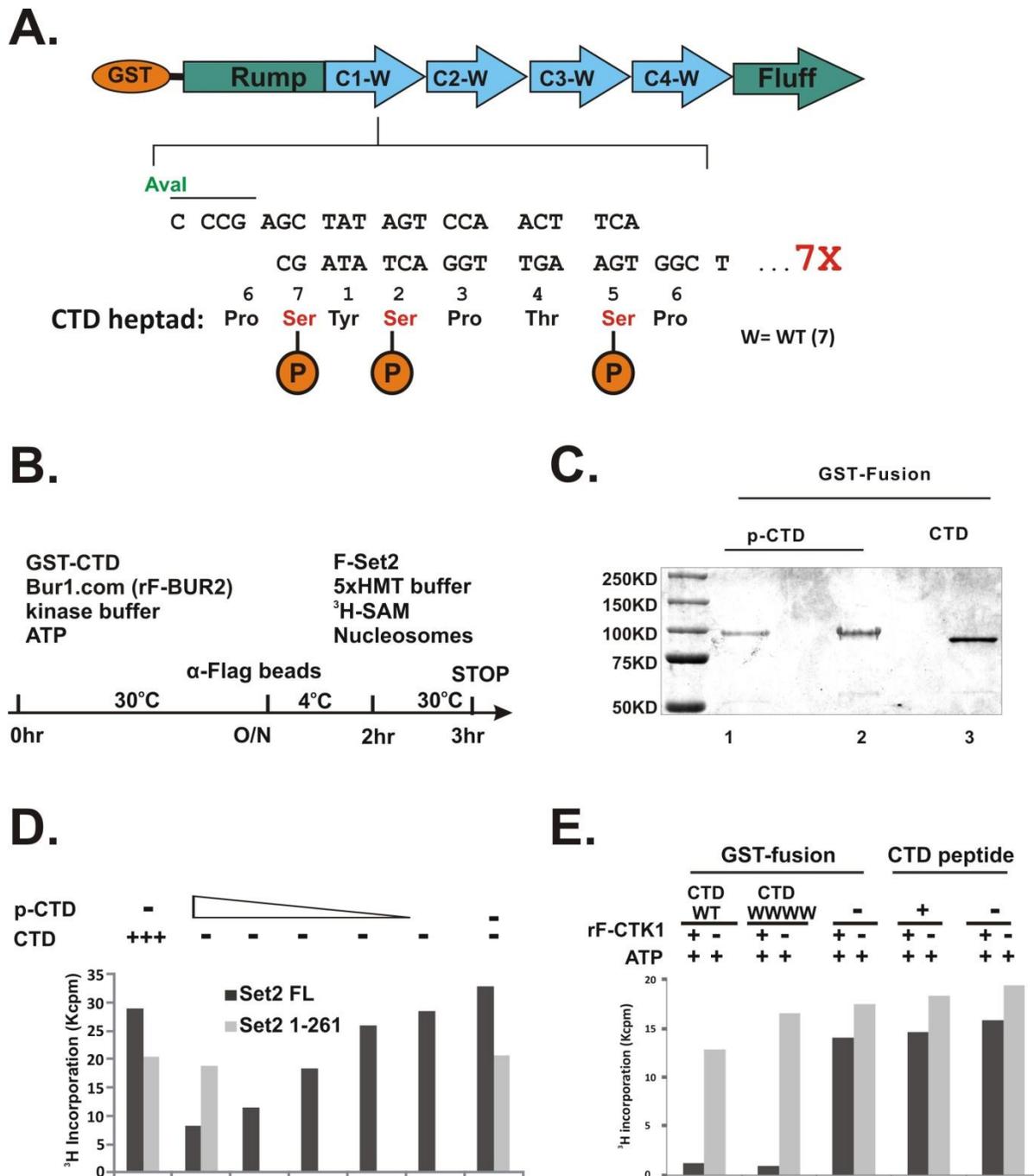


**Figure 3.3 Ser2 Phosphorylated CTD peptides (4X) do not inhibit Set2 activity.** HMT assay using synthesized four-heptad repeat CTD peptides: un-phosphorylated (CTD) or phosphorylated on Ser2 (S2P-CTD) to test the inhibitory effect upon F-Set2. Fixed amount of Set2 was incubated with 5-fold serial titrations of CTD peptides, then subjected to HMT assay using nucleosome substrates



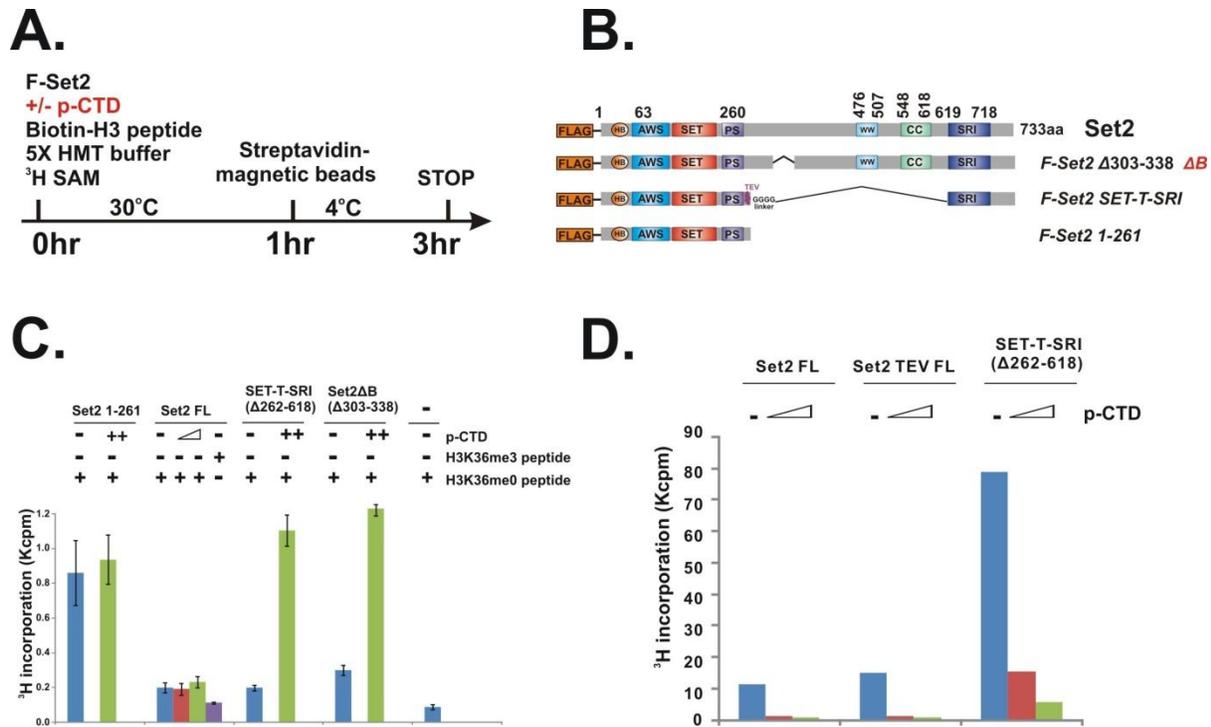
**Figure 3.4 A clone strategy to engineer mutations at any position along CTD.** (A) The scheme of pBSaK vector (pBL766) generated by mutating all the BtsI and NgoMIV restriction sites on pBluescript (I) KS plasmid. (B) The scheme of individual oligo adaptors inserted into pBSaK to form “CTD carrier” vectors C1 to C4. (C) The scheme illustrating the seven-heptad repeats of wild type (“W”, left) or S2, 5A (red “1”, right) mutant on “CTD carrier” vector. The nucleotide

and amino acid sequences of single heptad are shown. *Ava*I restriction site is marked in green. The circled “P” indicates phosphorylation while the circled “X” indicates lack of the phosphorylation sites on mutant. (D) The scheme displaying the intermediate FL WT CTD tail construct pKSn-WWWW (pBL791) which contains four blocks of seven each wild type CTD heptad repeats (C1-W to C4-W) and is made by the ligation of each seven-heptad repeat fragment from “CTD carrier” vectors through unique restriction sites as shown on the top of the graph. The final CTD tail contains 28 repeats of CTD heptad in total as well as the non-repetitive regions of the natural CTD at both N-terminus and C-terminus which are named as “Rump” and “Fluff” respectively. (E) The glossary of CTD nomenclature: Letters and numbers in black indicate wild type while red indicates mutant. The left panel lists the building units of CTD heptad: “WT” indicates individual wild type CTD heptad as shown in (C) while “W” represents seven repeats of “WT” heptad. The “1\*” indicates individual mutant CTD heptad with Ser2,5 replaced by Ala as shown in (C) and “1” means seven repeats of S2,5A heptad. The middle panel lists the CTD tails used in this study that carry various combinations of wild type and mutant heptads. “W” “WT”, “1” and “1\*” are consistent with the left panel and “C1” indicates the manipulation of the first block of seven-heptad repeat as shown in (D). The bars at right panel illustrate the copy number and location of either wild type (black) or mutant (red) heptad according to each CTD variants listed in the middle panel.

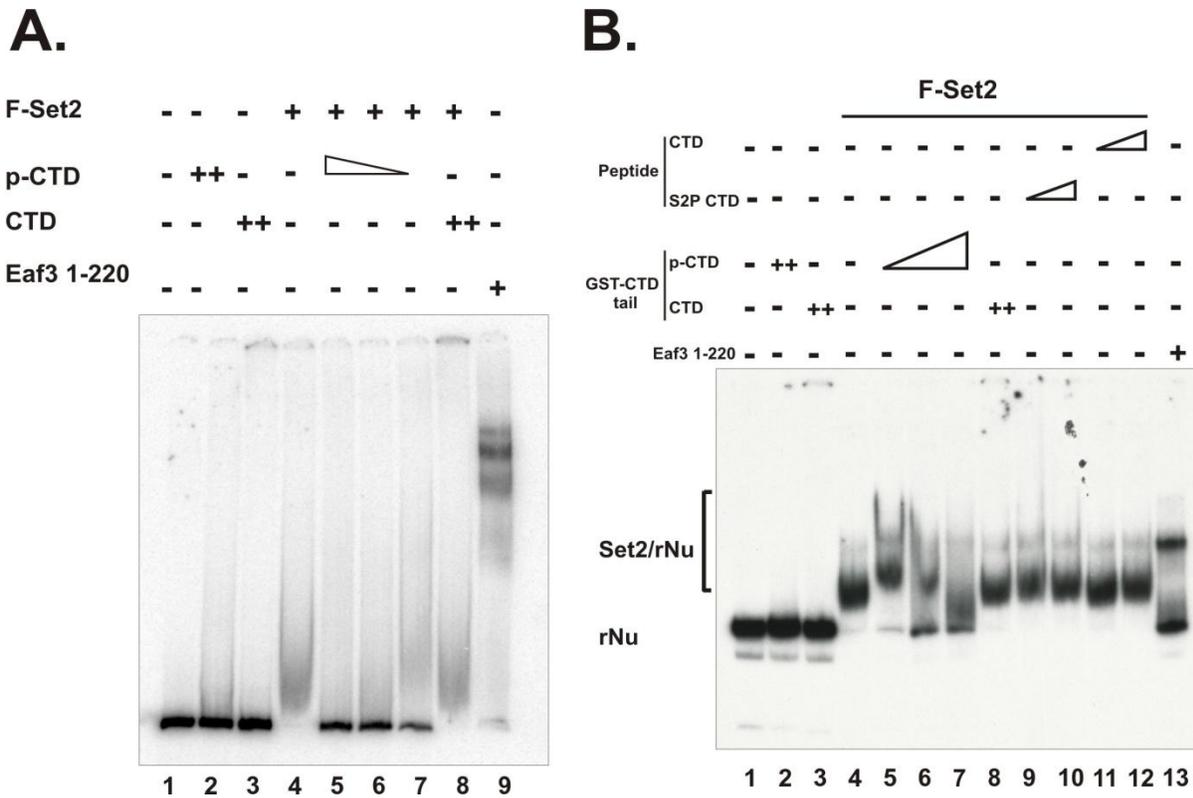


**Figure 3.5 Phosphorylated GST-CTD inhibits Set2 activity through the SRI domain.** (A) A scheme of engineered N-terminal GST-tagged FL WT CTD tail. The detailed structure of WT CTD tail has been shown and described in Fig 3.4. (B) The experimental scheme of checking the inhibitory effect of *in vitro* phosphorylated GST-CTD on Set2 activity. Baculovirus expressed Bur1 complex was purified through FLAG-tagged Bur2 subunit (F-Bur2) and used to

phosphorylate GST-CTD. After kinase reaction, anti-FLAG beads were applied to remove most of kinase complex from solution. The phosphorylated GST-CTD was added into HMT reaction containing HMT buffer, HeLa LON nucleosome, F-Set2 and  $^3\text{H}$  labeled SAM. (C) Coomassie blue staining of phosphorylated GST-CTD (p-CTD, Lane 1 and 2) compared with un-phosphorylated GST-CTD (CTD, Lane 3). (D) HMT assay showing p-CTD inhibits enzyme activities of recombinant Set2 (FL) but not Set2 1-261. (E) Comparing the inhibitory effects between GST-CTD tails and CTD peptide on Set2 activity in the assay coupling phosphorylation and HMT reactions. GST-CTD tails and CTD peptide were incubated with rF-CTK1 and ATP for one hour first and HeLa LON nucleosome, F-Set2 and  $^3\text{H}$  labeled SAM were immediately added for one more hour incubation to perform HMT reaction.



**Figure 3.6 Binding to phosphorylated GST-CTD neutralizes all SRI functions.** (A) An experimental scheme to test how phosphorylated GST-CTD tail influences the activities of Set2 variants on un-methylated H3 peptide. (B) A scheme illustrating FLAG-tagged Set2 variants used in (A). (C) The activity of Set2 middle domain deletion is rescued by p-CTD tail but not FL on H3 peptide. *In vitro* HMT assay as shown in (A) displaying activity changes of Set2 variants on synthesized H3 peptide upon adding *in vitro* phosphorylated GST-CTD tail. After reaction, the methylated H3 peptide was enriched on streptavidin-coated magnetic beads, and incorporated <sup>3</sup>H was measured. HMT results are represented as mean ± SEM, n=3. (D) p-CTD partially inhibits Set2 middle domain deletion activity on nucleosome. HMT assay using HeLa LON nucleosomes as substrate to test the inhibitory effect of phosphorylated RNA Pol II upon Set2 middle domain deletion (SET-T-SRI).

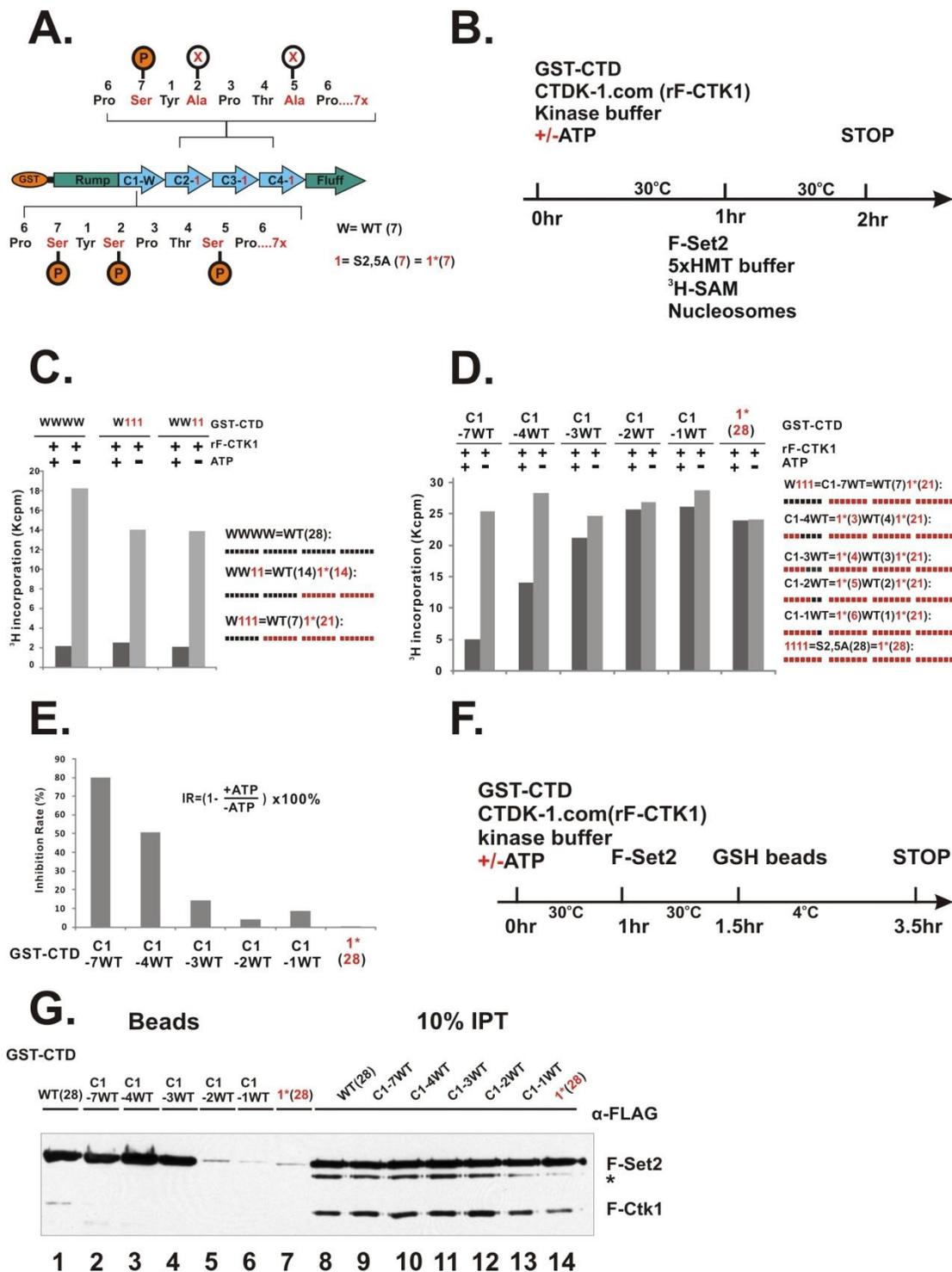


**Figure 3.7 Phosphorylated GST-CTD inhibits the binding of Set2 to DNA and nucleosomes.**

(A) EMSA assay showing F-Set2 binds to  $^{32}\text{P}$ -labeled DNA probes and phosphorylated CTD tail (p-CTD) competes with Set2/DNA interaction (Lane 4 compared with Lane 5, 6, 7). GST-Eaf3 1-220 protein served as positive control of proper DNA binding. (B) EMSA assay showing phosphorylated CTD tail (p-CTD), but not four-heptad repeat Ser2P CTD peptide (S2P CTD), prevents F-Set2 from binding to reconstituted mono-nucleosome (Lane 5, 6 and 7 compared with Lane 9 and 10). GST-Eaf3 1-220 protein served as positive control of proper nucleosome binding.



normalizing deuterium incorporation rate of p-CTD tail bound Set2 with that of Set2 alone. The resulted change ranges from -50% (blue) to +50% (red). Higher value suggests faster H/D exchange and more solvent accessibility at the region. The predicted yeast SRI/ DNA interacting residues (cyan *cross*) by Bind N (<http://bioinfo.ggc.org/bindn/>) web server and previously published human (purple *cross*) or yeast (green *cross*) Set2 SRI/ p-CTD peptides interacting residues are also indicated above the amino acid sequence. (B) EMSA assay showing that SRI is responsible for the binding between Set2 and double-strand DNA. End-labeled 216bp DNA that contains a 601 positioning sequence was used. The truncated Eaf3 proteins were served as a positive control for proper DNA binding. (C) Mapping of the residues shown in (A) on the yeast SRI domain structure (PDB 2C5Z (Venjnic et al 2006)) using PyMOL. The surface displays of following SRI residues are shown from four different angles (from top to bottom) when the SRI structure model is rotated: 1. Set2/p-CTD FL tail interacting residues from H/D exchange analysis (left panel blue gradient and pink regions), 2. The predicted SRI/DNA interacting residues (middle panel, cyan regions), and 3. the published p-CTD peptide binding residues of either human or yeast Set2 SRI domain (right panel, green for ySRI and purple for hSRI). The overlapped regions of Set2/p-CTD tail and predicted SRI/DNA interaction are highlighted by yellow dash circle.



**Figure 3.9 Extra copies of phosphorylated heptapeptide CTD inhibits Set2 HMT activity.** (A) A scheme of engineered GST-tagged FL CTD tail “W111” mutant. It contains seven-heptad wild type repeats (“W”) at the first block (C1) of CTD tail and seven-heptad S2, 5 A repeats (“1”) at

the rest of three blocks (C2-C4). (B) The experimental scheme to test the inhibitory effects of CTD variants on Set2 activity right after being phosphorylated by baculovirus expressed FLAG-tagged CTDK-1 complex (rF-CTK1). The GST-CTD tail variants and rF-CTK1 were first incubated with/without adding ATP for 1 hour at 30°C. Then F-Set2, HeLa nucleosomes, and <sup>3</sup>H labeled SAM were directly added and incubated for 1 more hour to perform HMT reactions. (C) Phosphorylation-HMT coupled assays illustrated in (B) showing the phosphorylation of seven or more wild type CTD heptads is sufficient to inhibit Set2 activity. The CTD nomenclatures are illustrated either by equations showing the number (“WT” for wild type and “1\*” for the S2, 5 A heptad) or bar presentations showing the locations (black for wild type and red for the mutant) of heptad units on CTD tails. (D) The same assay illustrated in (B) showing that further reducing WT CTD heptad on FL CTD tail rescues Set2 activity. (E) The relative inhibition of phosphorylated CTD tail variants converted from (D) by displaying the percentages of Set2 activity reduction upon ATP addition. (F) The experimental scheme of testing the interaction between F-Set2 and p-CTD tail variants: GST-CTD tail variants were incubated with rF-CTK1 with/without adding ATP at 30°C for 1 hour. Right after kinase reaction, F-Set2 was added and further incubated for a half hour. The glutathione coated beads (GSH beads) were added and the reaction was transferred to 4°C for two more hours of incubation. The eluate was resolved in 8% SDS-PAGE and anti-FLAG antibody was applied to detect CTD bound F-Set2. (G) Glutathione beads pull-down assay illustrated in (F) showing the FL CTD tail mutants containing three WT heptad repeats are required for Set2 binding.

## Materials and Methods

### Construction of plasmids and yeast Strains:

To construct GST-tagged Set2 and its N-terminal truncation mutants, FL and C-terminal portions of *SET2* fragments were generated by standard PCR reaction. The 5' primers for all Set2 fragments contain BamHI cutting sites. The 3' primer for FL Set2 contains XhoI cutting sites while others contain NotI cutting sites right after the stop codon. The PCR fragments were first cloned into PCR blunt vector (Life Technologies) and then FL Set2 was sub-cloned into pRET-GST N2 vector (pBL196) while mutants were sub-cloned into pRET-GST-X/N (pBL869) vector, which was modified from pRET-GST N2 vector by inserting XhoI/NotI sites.

To generate FLAG-tagged Set2 and Set2 1-261, the DNA fragments were directly made by PCR reaction using 5' primer containing XhoI and 3' primer containing NotI sites from yeast genomic DNA. For the Set2  $\Delta B$  and *SET-T-SRI* mutants, two parts of DNA fragments were prepared separately as illustrated in Figure 5B. One part contains the SET domain of Set2 flanked by the 5' XhoI site and the 3' end BamHI site (Set2 SET-T-SRI SET domain also contains a TEV-recognition site and a flexible linker). The other part includes the rest of Set2 C-terminal portion with 5' BamHI site and 3' NotI site. Two parts of fragments were cloned into XhoI/NotI digested pBP-HFT vector (pBL532) by three-way ligation.

To generate baculovirus expressing vectors for each individual subunit of Pol II CTD kinase complexes CTDK-1 and Bur1, all the ORFs flanked by XhoI/NotI sites were obtained by standard PCR reactions and first cloned into PCR Blunt vectors (Invitrogen). The fragments of *CTK1* and *BUR2* were cloned into pBacPAK-8-N-HisFlag-TEVsite (pBL532) through XhoI/NotI restriction sites to form pBP-HFT-Ctk1 (pWY092) and pBP-HFT-Bur2 (pWY095). Meanwhile the fragments of *CTK2*, *CTK3* and *BUR1* were cloned into pBacPAK8 (pBL541) to form pBP-Ctk2 (pWY093), pBP-Ctk3 (pWY094) and pBP-Bur1 (pWY096). pYIA-CTD-fusion-Set2 (KpnI/SnaBI) plasmid (Rpb1-Set2 fusion, pWY043) was constructed as described before (Chapter 2) and the resulting Rpb1-Set2 fusion shuffle strain with TAP-tagged Rpb3 subunit (YYW138) was generated by using the standard 5'-FOA counter selection method.

### **Construction of CTD plasmids:**

To construct the yeast Pol II CTD, we developed a strategy to build an engineered CTD tail containing four blocks of seven each CTD heptad repeats (the total number is 28 compared with 26 in natural yeast CTD). The final CTD construct also contains non-repetitive sequences of natural CTD at both N-terminus and C-terminus, which are named as “Rump” and “Fluff”, respectively. Our design allows us to make combinations of wild type and mutant (ex. S2A or S2, 5A) heptad at each individual block so that we are able to manage the number and location of wild type heptad on CTD tail.

We first mutated all the BtsI and NgoMIV restriction sites on pBluscript (I) KS (pBL194) vector by PCR-based mutagenesis to create the intermediate vector pBSak as shown in Fig 3.4 A (pBSKO-NgoM4-Bts1-2-3,pBL766). To make “CTD carrier” vectors pBSak-C1 to pBSak-C4 for each individual block of seven-heptad repeat (pBL767, 768,769 and 770), the different adaptors containing KpnI-XX-AvaI-XXX-SacI sequences (XX, XXX indicates unique restriction sites as shown in Fig 3.4 B) were made by oligo annealing and inserted into pBSak vector through KpnI/SacI sites. To construct CTD heptad fragment flanking with AvaI sites, the WT (P1258/1259) or mutant (S2, 5A in this study (P1352/1353)) heptad oligo fragment was synthesized and annealed. The annealed fragments were applied to tandem ligation through AvaI site within each pBSak-C1 to pBSak-C4 vector, respectively. The ligation products that only contain seven-heptad repeat were screened out and sequencing confirmed as shown in Fig 3.4 C. To this end, the seven-WT or seven-S2, 5A mutant-heptad repeat were constructed into pBSak-C1 to C4 vector.

Meanwhile the CTD “mothership” vector pKSn (pBSaK-Rpb1-KSn-Linker, pBL773) containing the “Rump” flanked with a KpnI site at 5’, a linker region containing BamHI-BtsI-NotI-AflIII-NgoMIV multi-cloning sites and the “Fluff” flanked with a SnaBI site at 3’ was also constructed upon the backbone of pBSak (pBL766). Thus individual WT or mutant seven-heptad fragments from each of the pBSak “CTD carrier” can be inserted into the

“mothership” through respective restriction sites on multi-cloning region to form desired combinations of heptad repeats flanked by non-repetitive “Rump” and “Fluff” regions.

The intermediate FL engineered WT CTD tail construct pKSn-WWWW (pBL791) or FL S2,5A construct pKSn-1111 (pBL858) were cloned by sequential ligation of either WT or mutant seven-heptad fragments on pBSaK “carrier” C1-C4 vectors through unique restriction sites as shown in Fig 3.4 D. The WT/mutant hybrid CTD tails (ex.pKSn-W111, pBL860 shown in Fig 3.9 A) was made by replacing the specific block on pKSn-WWWW, pKSn-1111 with mutant or WT seven-heptad fragment respectively. The resulting CTD tail constructs were ligated into KpnI/SnaBI digested pRET-GST-KSn vector (pBL942) which was modified from pRET-GST N2 (pBL869) by inserting KpnI-SmaI-SnaBI adaptor.

To generate GST-CTD mutants that carry specific combinations of WT or S2, 5A heptad within the first block of seven-heptad CTD repeat (C1), the 1<sup>st</sup> round tandem ligation was performed using WT heptad oligo and AvaI cut pBSak-C1 vector (pBL767). The ligated products that contain one to four copies of WT heptad repeat were screened out and confirmed by sequencing, respectively. These intermediate products were cut again by AvaI and the 2<sup>nd</sup> round tandem ligation was carried out using S2, 5A mutant heptad oligo. The resulting ligation products that only contain seven copies of heptad in total were screened out and confirmed by sequencing. All C1 fragments carrying different combinations of WT/S2,5A heptad were generated by BamHI/BtsI digestion. Each resulting fragments was subjected to three-way ligation reaction with BtsI/NotI cut C2 fragments (from pBL787) carrying seven-heptad S2,5A CTD mutant and BamHI/NotI digested pKsn CTD-WW11 vector (pBL792), which contains another 14 copies of S2,5A heptad. The resulting WT CTD tail fragments were further digested by KpnI/SnaBI, and cloned into pRET-GST-KSn (KpnI-SmaI-SnaBI) vector (pBL942).

### **Protein Purification:**

*E coli system:* Two flasks of 500ml of BL21-CodonPlus *E coli* cells transformed with indicated plasmids were cultured in LB medium and supplemented with 200µg/ml of Ampicillin (Sigma)

for 2-3 hours until OD 600 reached 0.4-0.6. Cell culture was cooled down on ice for 2 minutes and 200 $\mu$ l of 0.5M Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG Fisher) was added in each flask to induce the protein expression at 16 °C for another 4 hours. The cells were harvested and washed by 30ml of cold STE buffer (0.1M NaCl, 10mM Tris-HCl pH 8.0 and 1mM EDTA pH8.0). The cell pellets were re-suspended with 10 ml of 2M NaCl GST buffer (2M NaCl,10mM Tris-HCl pH 8.0, 20mM NaPO<sub>4</sub> pH 6.8, 10mM  $\beta$ -Me and 0.01% NP40 freshly supplemented with 10 $\mu$ l of 1M Benzamidine-HCl, 30  $\mu$ l of 200mM PMSF and 100 $\mu$ l of 1M DTT) and sonicated for 15 seconds twice. Repeat sonication once after adding 1.5 ml 10% Triton X-100 into cell suspension. The sonicated cell lysates were incubated at 4°C for 30 minutes and centrifuged at 14'000 rpm for 10 minutes. The clarified lysates after centrifugation were incubated with 250 $\mu$ l of glutathione Sepharose 4B (GE Healthcare) at 4 °C for 4 hours to overnight. After intensive wash with 1 M and 0.1 M GST buffer, the protein bound glutathione beads were loaded on 10ml column (Bio-Rad) and equilibrated with about 3ml of GST elution buffer (50mM NaCl,25mM Tris-HCl pH 8.0, 10% Glycerol and 10mM  $\beta$ -ME ). The bound proteins were eluted with GST elution buffer containing 10mM glutathione (Sigma) and dialyzed in 1 litter GST elution buffer for 3 times and concentrated to about 1 $\mu$ g/ $\mu$ l.

*Insect cell system:* To purify FLAG-tagged Set2 variants 100ml Sf21 cells (1X10<sup>6</sup> cells/ml) supplemented with 10% FBS (Sigma), 1% Penicillin Streptomycin (Life Technologies) were infected with proximally 10ml of each P2 virus and incubated at 27°C for 48 hours. Cells were collected and washed with 10ml 1X PBS buffer. Cell pellets were then lysed in 10ml BV-lysis buffer (50mM HEPES PH7.9, 500mM NaCl, 10% glycerol,0.5mM EDTA, 2mM MgCl<sub>2</sub> and 0.2% Triton X-100) on ice for 30 minutes. The lysates were clarified by ultracentrifugation (Beckman 50.2 Ti 40,000 rpm) for 30 minutes. 200  $\mu$ l of anti-FLAG M2 affinity beads (Sigma) were added to the supernatants and incubated at 4 °C for at least 2 hours. Beads were then washed with 10ml BV-lysis buffer for 3 times and eluted with 600 $\mu$ l of 500 $\mu$ g/ml of 3xFLAG peptides (Sigma

F4799) twice. The eluents were concentrated to about 150 $\mu$ l using 10kd cut-off concentrators (Amicon).

To purify CTD kinase complex CTDK-1(F-Ctk1) and Bur1 (F-Bur1) through FLAG-tagged Ctk1 and Bur2 subunit respectively, equal amount of P2 virus (10ml in total) of individual subunit was added into 100ml Sf21 cells ( $1 \times 10^6$  cells/ml) supplemented with 10% FBS (Sigma), 1% Penicillin Streptomycin (Life Technologies). Similar purification protocol was applied as described above with the exception of using the BV-lysis buffer containing 300mM instead of 500mM NaCl (50mM HEPES PH7.9, 300mM NaCl, 10% glycerol, 0.5mM EDTA, 2mM MgCl<sub>2</sub> and 0.2% Triton X-100) for cell lysis.

TAP purification from yeast: The TAP purification was performed essentially as described in previous publication (Li et al., 2003b). Briefly, 6 liters of TAP-tagged Set2 derivative strains were grown in YPAD media supplemented with 2% Dextrose at 30°C OD<sub>600</sub>=2-3. The cell pellets were re-suspended in Extraction buffer (E buffer) (40mM HEPES pH 7.5; 350mM NaCl; 10% Glycerol; 0.1% Tween 20) supplemented with Protease Inhibitor Cocktail (Sigma P8215 ) and disrupted through bead-beating (Biospec). Crude whole cell extracts were clarified through ultracentrifugation (Beckman 50.2 Ti at 45,000 rpm) and then incubated with 600 $\mu$ l IgG-sepharose (GE) at 4 °C for overnight. Bound proteins were cleaved from the beads by 100 units of TEV protease (Invitrogen) at 18 °C for 3hours. The eluent was mixed with 3ml of Calmodulin Binding Buffer (10mM Tris.HCl pH8.0, 150mM NaCl, 1mM Magnesium acetate, 1mM Imidazole, 2mM CaCl<sub>2</sub>, 10mM  $\beta$ -ME, 0.1% NP40, 10% Glycerol) and 3 $\mu$ l of 1M CaCl<sub>2</sub>, and subsequently incubated with 600 $\mu$ l Calmodulin-Sepharose (GE) at 4 °C for one hour. The bead bound proteins were finally eluted by Calmodulin Elution Buffer (10mM Tris.HCl pH8.0, 150mM NaCl, 1mM Magnesium acetate, 1mM Imidazole, 2mM EGTA pH8.0, 10mM  $\beta$ -ME, 0.1% NP40, 10% Glycerol) and concentrated using Amicon concentrators.

#### **CTD peptide pull-down Assay:**

Biotinylated peptides that corresponding to 4 heptad repeats of unphosphorylated (CTD), Serine2 phosphorylated (Ser2P) or Serine 5 phosphorylated (Ser5P) of Pol II CTD were immobilized on Dynabeads® M-280 Streptavidin (Invitrogen) at 4 °C for 2 hours. The beads carrying 8 µg of each peptide were then used for pull-down reaction with 1µg FL Set2 or equal molar amount of mutant proteins in Peptide Binding Buffer (25mM Tris.HCl pH 8.0, 50mM NaCl, 5% Glycerol, 0.03% NP40 and 1mM DTT). The bound proteins were eluted with 15ml of 3XSDS buffer at RT for 15 minutes and analyzed by Western blotting analysis using anti-GST (Santa Cruz) antibodies.

#### **Preparation of whole-cell extracts for immunoblotting:**

Cells were grown in 3ml YPD or synthetic drop-out media supplemented with 2% dextrose or 2% Galactose at 30°C overnight, until OD reached 1-1.2. Cell pellets were re-suspended in 45µl of STE buffer (containing 500mM NaCl, 10mM Tris HCl PH8.0 and 1mM EDTA) and 40µl of 3xSDS buffer. 100µl of 0.5mm glass beads were then added and boiled at 95°C for 5 minutes. Samples were then vigorously vortexed to break the cells, and cell suspensions were clarified by spinning at 14000rpm for 5minutes. Anti-H3K36me3 (Abcam, 9050), anti-H3K36me2 (Abcam, 9049), anti-H4 (Abcam, 10158), and anti-FLAG HRP (Sigma, A8592) antibodies were used according to manufacturers' instruction. For anti-H3K36me2 antibody, 5% BSA was used for blocking, and 1% BSA were presented in all antibody incubation.

#### **Preparation of phosphorylated Pol II and GST-CTD:**

To phosphorylate RNA Pol II complex, 100µg of Pol II purified through Rpb9-TAP was incubated with 30µg of F-Ctk1 complex in 2.6 ml reaction system, which contained 260µl of 10x p42 MAP Kinase buffer (500mM Tri HCl pH7.5, 100mM MgCl<sub>2</sub>, 1mM EDTA, 20mM DTT and 0.1% NP40)( NEB B6022) and 50µl of 100mM ATP (Sigma) at 30°C for 20 hours. After checking the completion of phosphorylation by silver staining, the kinase buffer was replaced by Calmodulin Binding Buffer (10mM Tris.HCl pH8.0, 150mM NaCl, 1mM Magnesium acetate,

1mM Imidazole, 3mM CaCl<sub>2</sub>, 10mM β-ME, 0.1% NP40, 10% Glycerol) by using 100kD molecular weight cutoff Amicon Ultra (Millipore). Phosphorylated Pol II complex was immobilized by 600μl of Calmodulin Sepharose 4B (GE Healthcare) to get rid of most of the kinase, and then eluted with Calmodulin Elution Buffer (10mM Tris.HCl pH8.0, 150mM NaCl, 1mM Magnesium acetate, 1mM Imidazole, 2mM EGTA pH8.0, 10mM β-ME, 0.1% NP40, 10% Glycerol) and concentrated.

To *in vitro* phosphorylate GST-CTD, 200μg of FL engineered GST-CTD tail (CTD-WWWW) was incubated with about 50μg of F-Bur1 complex in 1ml reaction system, which contains 1x p42 MAP Kinase buffer and 2mM ATP at 30°C for 20 hours. After kinase reaction, 20ul of anti-FLAG M2 affinity beads (Sigma) were added to immobilize the most of the kinase and the supernatant was collected and concentrated by 10kD molecular weight cutoff Amicon Ultra (Millipore). The final phosphorylated product (p-CTD) was subjected to 10% SDS-PAGE and stained by Coomassie blue.

### **Histone methyltransferase assays (HMT):**

Standard HMT reactions were carried out in a 20μl system using 1xHMT buffer (50 mM Tris.HCl pH 8.0, 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM DTT, 5 % Glycerol) supplemented with 0.5μl of radio-labeled <sup>3</sup>H S-Adenosyl methionine (SAM) (80 Ci/mmol, Perkin Elmer) at 30 °C for 1 hour (Li et al., 2003b). In phosphorylated Pol II inhibition assay, 0.1 μg of F-Set2 and 1μg of HeLa LON nucleosome were used, the initial amount of p-Pol II was 0.66μg and 3-fold serial titrated to 0.01μg. In FL p-CTD tail inhibition assay, 1μg of F-Set2 and equal molar of F-Set2 1-261 were used. The amount of FL p-CTD tail started with 1μg and 3 folds titrated to 0.01μg. Scintillation counting-based filter-binding assay was performed as described previously (Eberharter et al., 1998).

For the HMT assay using H3 peptides as substrates, 2μg of biotinylated H3 peptides (either un-methylated or H3K36me<sub>3</sub>), 1μg of F-Set2, and equal molar of mutants were used. After methyltransferase reaction, the salt concentration of the total reaction solution was increased to

1M by supplementing 5M NaCl. 5µl of Dynabeads® M-280 Streptavidin (Invitrogen) was added to immobilize H3 peptides at 4°C for 2 hours. The beads were washed by 500 µl of Peptide Binding Buffer (25mM Tris.HCl pH 8.0, 50mM NaCl, 5% Glycerol, 0.03% NP40 and 1mM DTT) for 3 times, and ScintiSafe™ 30% Cocktail (Fisher) was applied for scintillation count.

For the CTD inhibition assays using un-purified p-CTD products following kinase reactions, either 3µg of TAP purified Pol II-Set2 fusion complex or 1µg of GTS-CTD variants were first incubated with F-Ctk1 complex in a 10µl reaction containing 1µl of 10x p42 MAP Kinase buffer (NEB B6022) and 0.5 µl of 100mM ATP (Sigma) at 30 °C for 1 hour. Then 20µl of 5xHMT (Magnesium free) buffer (250 mM Tris.HCl pH 8.0, 250 mM NaCl, 10 mM DTT, 25 % Glycerol), 0.5ul of <sup>3</sup>H labeled S-Adenosyl methionine (80 Ci/mmol, Perkin Elmer), 1µg of HeLa nucleosome and 80ul of H<sub>2</sub>O were added for one more hour incubation. In the assay that FL p-CTD tail variants were used, 0.8µg of F-Set2 was added after kinase reaction as well.

#### **Recombinant histone purification and nucleosome reconstitution for EMSA assay:**

Xenopus recombinant histones (H3, H4, H2A, H2B) were individually expressed in BL21CodonPlus-RIL (Stratagene) cells and purified as described (Li et al., 2007b). Histone octamers were assembled and fractionated through gel-filtration column Superdex 200. To generate mono-nucleosomes with <sup>32</sup>P labeled 601 positioning DNA fragments, similar strategy was applied as previously described ( Yun et al, 2012) 20µl of 5' primer of 601 DNA sequence was end labeled with γ-<sup>32</sup>P labeled ATP (Perkin Elmer) by T4 PNK kinase (NEB) at 37°C for 1 hour. The labeled 5' primer was mixed with same amount of 3' primer and 10ng/µl of pBS-196L-1X plasmid (pBL630) to PCR amplify 601 DNA template using lab made Pfu DNA polymerase. The final PCR products were loaded onto a 2.0% agarose gel in 1.5× TAE buffer. The DNA bands with the correct sizes were excised and purified using standard Gel extraction Kits (Qiagen Gel). Nucleosome reconstitution was carried out by adding 5 M NaCl, 2 mg/ml BSA, H<sub>2</sub>O, 1 pmol of radio-labeled DNA, and histone octamers into a 10 µl reaction. While the theoretical mass ratio between octamer and DNA should be 1:1 while the practical ratio was

determined by the pilot titration assays. The final concentrations of NaCl and BSA were 2 M and 0.1 mg/ml, respectively. The salt concentration was gradually reduced by adding 3.3, 6.7, 5,3.6, 4.7, 6.7, 10, 30, and 20  $\mu$ l of the initial buffer ( 50mM HEPES Ph7.5, 1mM EDTA,0.1% NP40,20% Glycerol, 5mM DTT 0.5mM PMSF and 100 $\mu$ g/ml BSA)with a 15-min interval between each step while samples were incubated at 30°C.The reaction was brought to 0.1 M NaCl by adding 100  $\mu$ l of final buffer (10mM Tris HCl pH7.5, 1mM EDTA,0.1%NP40,20% Glycerol 5mM DTT, 0.5mM PMSF and 100 $\mu$ g/ml BSA ) and incubated at 30°C for another 15 min. The efficiency of nucleosome reconstitution was examined by running 2  $\mu$ l of each sample on a 5% native PAGE gel in 0.3 $\times$  TBE buffer. To purify reconstituted nucleosomes, all the samples were loaded onto 5% PAGE gel in 0.3 $\times$ TBE running at 11 V/cm at 4°C for 5 hour. Cover the wet gel with plastic wrap and expose directly to an X-ray film for 30 minutes. According to the signal on X-ray film the gel pieces right at the position where nucleosomes migrated were cut by razor blade and smashed by pipetting up and down. Approximate 1.5x gel volume of Gel Elution buffer (~200–400  $\mu$ l) (10mM Tris HCl pH7.4, 100mM NaCl, 1mM EDTA, 5mM DTT,0.5mM PMSF and 0.1mg/ml BSA ) was added to elute the nucleosomes and the sample was placed at 4°C on a roller mixer for overnight incubation. The next day after 21,000  $\times$  g centrifuge at 4°C for 5 minutes, the barrel tips were used to withdraw the liquid from the gel and then transfer it to a 0.5-ml Eppendorf tube.  $\frac{1}{4}$  volume of 5 $\times$  Gel Final buffer (10mM Tris-HCl pH7.4, 100mM NaCl, 1mM EDTA, 50% Glycerol 0.25% NP40, 5mM DTT,0.5mM PMSF and 0.1mg/ml BSA)was added to the eluates and the resulting nucleosome solution was stored at 4°C. 1.5 $\mu$ l of nucleosome sample was taken for the scintillation counting.

### **Gel shift assays (EMSA):**

EMSA reactions were carried out in a 15- $\mu$ l system, containing 10 mM HEPES pH 7.8, 50 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM DTT, 0.25 mg/ml BSA, 5% Glycerol and 0.1 mM PMSF (1  $\times$  EMSA buffer). The samples were subjected to 4% PAGE gel running at 4 °C. In Set2/DNA binding assay, 1 $\mu$ g of F-Set2 was firstly incubated with 2000cpm scintillation count <sup>32</sup>P labeled

216 bp 601 DNA fragments at 30°C for 30 minutes, then either phosphorylated or un-phosphorylated GST-CTD was added into reaction for 30 more minutes of incubation. The amount of GST-CTD was started at 0.3µg and serially diluted to 0.033µg by 3-fold.

For the EMSA assay testing that p-CTD competes with Set2/nucleosome binding, 1µg of Set2 was used to pre-incubate with <sup>32</sup>P labeled 196bp 601 DNA nucleosomes at 30°C for 30 minutes. Then 0.5, 0.167 and 0.056µg of either phosphorylated or un-phosphorylated GST-CTD or 0.2 and 0.1µg of four-heptad CTD repeat peptide (un-phosphorylated (CTD) or phosphorylated on Ser2 (S2P)) was added into reaction respectively for 30 more minutes of incubation.

#### **GST-CTD pull-down assays:**

1.5µg of GST-CTD variants were incubated with 2.25µg of F-Ctk1 complex in 10µl reaction systems containing 1x p42 MAP Kinase buffer (NEB) with/without 10mM of ATP (Sigma) at 30°C for 1 hour. Then 20µl of 5xHMT (Magnesium free) buffer (250 mM Tris.HCl pH 8.0, 250 mM NaCl, 10 mM DTT, 25 % Glycerol), 80ul of H<sub>2</sub>O and 0.8 µg of F-Set2 were added for a half more hour incubation. 10µl of glutathione sepharose 4B (GE) 50/50 slurry which was pre-equilibrated with 1x HMT buffer was added, then the whole reaction was incubated at 4°C for 2 more hours. The glutathione beads were washed 3 times by 150mM GST NaCl buffer (150mM NaCl, 10mM Tris-HCl pH 8.0, 20mM NaPO<sub>4</sub> pH 6.8, 10mM β-Me and 0.01% NP40) and boiled with 10µl of 3xSDS buffer for 5 minutes. The elutes were resolved in 10% SDS-PAGE, and Western blotting analysis was performed using anti-FLAG HRP antibody (Sigma, A8592).

Table 3.1 Plasmid List

| <b>Plasmid</b> | <b>Backbone</b>    | <b>Description</b>  | <b>Source</b> |
|----------------|--------------------|---|---------------|
| pWY001/A       | pBL196             | <i>pRET-GST-Set2 (BamHI/XhoI)</i>                                     | this study    |
| pWY128         | pBL869             | <i>pRET-GST-Set2-SRI (BamHI/NotI)</i>                                 | this study    |
| pWY139         | pBL869             | <i>pRET-GST-Set2-477C (BamHI/NotI)</i>                                | this study    |
| pWY141         | pBL869             | <i>pRET-GST-Set2-262C (BamHI/NotI)</i>                                | this study    |
| pWY005         | pBL532             | <i>pBP-HFT-Set2</i>   | this study    |
| pWY037         | pBL532             | <i>pBP-HFT-Set2 1-261</i>   | this study    |
| pWY061         | pBL532             | <i>pBP-HFT-Set2 SET-G5-SRI TEV (<math>\Delta</math>262-618)</i>       | this study    |
| pWY105         | pBL532             | <i>pBP-HFT-Set2 <math>\Delta</math>B (<math>\Delta</math>303-338)</i> | this study    |
| pWY043         | pBL855             | <i>pYIA-CTD-fusion-Set2 (KpnI/SnaBI)</i>                              | this study    |
| pWY044         | pBL855             | <i>pYIA-CTD-fusion-Set2 1-618 (KpnI/SnaBI)</i>                        | this study    |
| pWY045         | pBL855             | <i>pYIA-CTD-fusion-Set2 1-261 (KpnI/SnaBI)</i>                        | this study    |
| pWY046         | pBL855             | <i>pYIA-CTD-fusion(KpnI/SnaBI)</i>                                    | this study    |
| pBL766         | pBluescript (I) KS | <i>pBSKO-NgoM4-Bts1-2-3</i>   | Li Lab        |
| pBL767         | pBL766             | <i>pBSaK-C1</i>   | Li Lab        |
| pBL768         | pBL766             | <i>pBSaK-C2</i>   | Li Lab        |
| pBL769         | pBL766             | <i>pBSaK-C3</i>   | Li Lab        |
| pBL770         | pBL766             | <i>pBSaK-C4</i>   | Li Lab        |
| pBL787         | pBL768             | <i>pBSaK-C2-S25A-7CTD</i>   | Li Lab        |
| pWY227         | pBL767             | <i>pBSaK-C1-4WT</i>   | this study    |
| pWY228         | pBL767             | <i>pBSaK-C1-3WT</i>   | this study    |
| pWY229         | pBL767             | <i>pBSaK-C1-2WT</i>   | this study    |
| pWY230         | pBL767             | <i>pBSaK-C1-1WT</i>   | this study    |
| pBL773         | pBL766             | <i>pBSaK-Rpb1-KSn-Linker</i>  | Li Lab        |
| pBL791         | pBL773             | <i>pKSn-WWWW</i>  | Li Lab        |
| pBL858         | pBL773             | <i>pKSn-1111</i>  | Li Lab        |
| pBL860         | pBL773             | <i>pKSn-W111</i>  | Li Lab        |
| pWY231         | pBL773             | <i>pKSn-C1-4WT</i>  | this study    |
| pWY232         | pBL773             | <i>pKSn-C1-3WT</i>  | this study    |
| pWY233         | pBL773             | <i>pKSn-C1-2WT</i>  | this study    |
| pWY234         | pBL773             | <i>pKSn-C1-1WT</i>  | this study    |
| pBL942         |                    | <i>pRET-GST-KSn (KpnI-SmaI-SnaBI)</i>                                 | Li Lab        |
| PHC033         | pBL942             | <i>pRET-GST-CTD-WWWW</i>  | this study    |
| PHC031         | pBL942             | <i>pRET-GST-CTD-WT</i>  | this study    |
| pWY097         | pBL942             | <i>pRET-GST CTD-W111</i>  | this study    |
| pBL946         | pBL942             | <i>pRET-GST-KSn-1111</i>  | this study    |
| pWY223         | pBL942             | <i>pRET-GST-CTD C1-4WT</i>  | this study    |
| pWY224         | pBL942             | <i>pRET-GST-CTD C1-3WT</i>  | this study    |

|               |         |                                    |            |
|---------------|---------|------------------------------------|------------|
| <b>pWY225</b> | pBL942  | <i>pRET-GST-CTD C1-2WT</i>         | this study |
| <b>pWY226</b> | pBL942  | <i>pRET-GST-CTD C1-1WT</i>         | this study |
| <b>pHC210</b> | pBL869  | <i>pRET-GST-Eaf3 1-220</i>         | Li Lab     |
| <b>pWY092</b> | pBL532  | <i>pBP-HFT-Ctk1</i>                | Li Lab     |
| <b>pWY093</b> | pBL541  | <i>pBP-Ctk2</i>                    | Li Lab     |
| <b>pWY094</b> | pBL541  | <i>pBP-Ctk3</i>                    | Li Lab     |
| <b>pWY095</b> | pBL532  | <i>pBP-HFT-Bur2</i>                | Li Lab     |
| <b>pWY096</b> | pBL541  | <i>pBP-Bur1</i>                    | Li Lab     |
| <b>pBL541</b> |         | <i>pBacPAK8</i>                    | Li Lab     |
| <b>pBL532</b> |         | <i>pBacPAK-8-N-HisFlag-TEVsite</i> | Li Lab     |
| <b>pBL630</b> | pBS-201 | <i>pBS-196L-1X</i>                 | Li Lab     |

Table 3.2 Strain List

| Name   | Parental strain | Genotype  | Source     |
|--------|-----------------|---|------------|
| YYW037 | PH499           | <i>MATa ura3-52 lys2-801 ade2-101 his3-Δ200 leu2-Δ1 Set2-C-TAP::TRP1</i>  | Li Lab     |
| YYW040 | BY4742          | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 Set2 1-618-C-TAP::URA3</i>   | this study |
| YBL360 | BY4741          | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Rpb9-TAP::HIS3</i>   | Li Lab     |
| YYW114 | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY046(pYIA CTD-fusionKpnI/SnaBI Leu2)</i>  | this study |
| YYW115 | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY045(pYIA CTD-fusion-Set2 1-261KpnI/SnaBI Leu2)</i>   | this study |
| YYW117 | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY046(pYIA CTD-fusion KpnI/SnaBI Leu2)</i>   | this study |
| YYW118 | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY045(pYIA CTD-fusion-Set2 1-261KpnI/SnaBI Leu2)</i>                                       | this study |
| YYW120 | Z26             | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pRP112 (RPB1 CEN URA3)</i>  | this study |
| YYW121 | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3ΔSET2::HPH + pRP112 (RPB1 CEN URA3 )</i>  | this study |
| YYW123 | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY044(pYIA CTD-fusion-Set2 1-618 KpnI/SnaBI Leu2)</i>  | this study |
| YYW124 | YYW120          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 +pWY043(pYIA CTD-fusion-Set2 KpnI/SnaBI Leu2)</i>  | this study |
| YYW125 | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY044(pYIA CTD-fusion-Set2 1-618 KpnI/SnaBI Leu2)</i>                                      | this study |
| YYW126 | YYW121          | <i>MATalpha hisΔ200 ura3-52 leu2-3,112 rpb1Δ187:: HIS3 ΔSET2::HPH+pWY043(pYIA CTD-fusion-Set2 KpnI/SnaBI Leu2)</i>  | this study |
| YYW122 | CKY283          | <i>MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128Δ gal10Δ56 rpb1Δ::CLONATMX Rpb3-TAP::KlacTRP1+pRP112 (RPB1 CEN URA3)</i>                        | this study |
| YYW138 | YYW122          | <i>MATa ura3-52 his3Δ200 leu2Δ1 or Δ0 trp1Δ63 met15Δ0 lys2-128Δ gal10Δ56 rpb1Δ::CLONATMX Rpb3-TAP::KlacTRP1+pWY043 (pYIA CTD-fusion-Set2 KpnI/SnaBI Leu2)</i> | this study |

## Chapter 4 THE SPECIFIC STATES OF H3K36 METHYLATION REGULATES DIFFERENT CHROMATIN FUNCTIONS

### Abstract

Posttranslational modifications (PTMs) of histones, such as acetylation, methylation, phosphorylation, and ubiquitylation play important roles in regulating chromatin functions. Mono-, di- or tri- methylation of lysine residues further increases the complicity of histone PTM signaling. The F/Y switch of histone methyltransferases has been shown to control catalysis and products specificities. To characterize the methyl-state specific functions of H3K36me, we generated two F/Y switch mutants. The Y149F mutant only confers K36me3 but no K36me1 and K36me2, whereas the F234Y mutant can only methylate up to the state of K36me2 both *in vitro* and *in vivo*. These two mutations cause distinct genetic interactions with histone chaperone Spt16 and the CTD proline isomerase Ess1, suggesting that they play different roles during transcription. Moreover, we found that H3K36me2 but not K36me3 was required for the non-homologous end joining DNA repair pathway. Collectively, our results demonstrate that different states of H3K36me have distinct signaling functions in cells.

## Introduction

### H3K36 methyltransferases in yeast and higher eukaryotes

Most histone methyltransferases that have been identified so far contain a conserved catalytic SET domain except the Dot1 family which specifically methylates H3K79 (Feng et al., 2002; van Leeuwen et al., 2002). In yeast, Set2 is the only methyltransferase performing all three methylation reactions at H3K36 (Strahl et al., 2002), while in higher eukaryotes, multiple H3K36 methyltransferase have been discovered and each has its own product specificities. For instance, Setd2/Hypb, the ortholog of yeast Set2, is thought to be the only H3K36 tri-methylase (Edmunds et al., 2008; Yuan et al., 2009) suggesting the unique function of H3K36me<sub>3</sub>. Meanwhile Nuclear receptor SET Domain-containing (NSD) family members (NSD1/2/3) have been found as H3K36 mono- and di- methyltransferases both *in vivo* and *in vitro* (Kuo et al., 2011; Li et al., 2009b; Rayasam et al., 2003; Yang et al., 2008). Moreover, other H3K36 methyltransferases are also identified with similar product specificities including ASH1L (An et al., 2011; Gregory et al., 2007), SMYD2 ((Abu-Farha et al., 2008; Brown et al., 2006), SETMAR (Fnu et al., 2011; Lee et al., 2005) and SETD3 (Eom et al., 2011). Collectively it suggests that different state of methylation may have different functions in cells.

Yeast Set2 and its orthologs human Setd2/Hypb and fly Hypb share highly conserved amino acid sequences and functional domains including the catalytic SET domain (Sun 2005), C-terminal Set2-Rpb1 interacting domain (SRI) and WW domain. SRI directly bind to phosphorylated RNA Pol II and help Set2 travel with elongation polymerase, (Bell et al., 2007; Kizer et al., 2005; Sun et al., 2005) while WW domain has been shown to bind human Huntington protein Htt (Gao et al., 2014). H3K36 methyltransferase Set2, Set2/Hypb, NSD family members and ASH1 are classified in the same family of SET containing methyltransferase because of the highly conserved amino acid sequence in SET domain (Dillon et al., 2005). The SET domain contains a series of  $\beta$  strands which are packed together with pre-SET (N-SET) and post-SET (C-SET) regions and fold into three discrete sheets that surrounding a knot like structure (Taylor et al., 2003). This unusual “pseudo-knot” is formed by threading the C-terminal

segment of the SET domain through a short loop of preceding stretch of the sequence. A narrow hydrophobic channel that links the substrate lysine or methylated lysine and cofactor product SAH has been observed in all known SET domain complex structures (Qian et al., 2006; Southall et al., 2009; Xiao et al., 2005; Xiao et al., 2003a; Zhang et al., 2003): The cofactor AdoMet and substrate bind to distinct clefts on the opposite surfaces of the SET domain, while methylated lysine is buried in the channel at the enzyme active site formed by several conserved hydrophobic residues (Qian and Zhou, 2006).

### **Product specificities of histone methyltransferases**

The lysines within histone can be methylated to three degrees (Mono-, di- and tri- methyl states). By comparing amino acid sequences between methyltransferases, a likely explanation for different product specificities was proposed that some residues in the lysine binding channel sterically exclude the target lysine side chain with methyl group(s) (Zhang et al., 2003).

There are two highly conserved motifs in the SET catalytic center: ELx(F/Y)DY and NHS/CxxPN which help to form an active site next to the methyl donor binding pocket and substrate peptide binding cleft (Qian and Zhou, 2006). The F/Y switch model within conserved ELx(F/Y)DY motif was first proposed by comparing the residues surrounding the target lysine in the structure of DIM-5 (a H3K9tri-methyltransferase) to those of SET7/9 ( a H3K4 mono-methyltransferase) (Zhang et al., 2003). Mutation of Phe of DIM-5 or human G9a to Tyr converts these two enzymes from tri- to mono- or di-methyltransferase (Collins et al., 2005). Likewise, switching the conserved Tyr to Phe alters product specificity of SET7/9 from mono- to di- /tri-methyl (Del Rizzo et al., 2010). It also has been shown in Set8, removing the hydroxyl group on Tyr in catalytic core stimulates tri-methyltransferase activity (Collins et al., 2005) because this alternation enlarges the diameter of the active site, accommodating the increasing size of the methylated  $\epsilon$ -amino group during methyl transfer reactions (Couture et al., 2008; Del Rizzo et al., 2010). In addition, yeast Set1 mutant Y1502F increased Set1 tri-methylase activity (Takahashi 2009) and Y334F mutant changes Set8 product specificity from a H4K20

mono-methyltransferase to a di-methyltransferase. (Couture et al., 2008). Collectively, these results shed lights on the mechanism of how SET domain governs the product specificities.

### **Functions of state-specific methylation.**

The combinatorial nature of histone amino-terminal modification reveals a “histone code” (Jenuwein and Allis, 2001). Each methylated lysine residue can exist in a mono-, di-, or tri-methylated state, further extending the complicity of this modification profile (Zhang et al., 2012c). Up to date, there are a handful of examples of how state specific methylation of a certain residue regulates distinct cellular processes.

*H3K4me1 and H3K4me3*: H3K4me1 and H3K4me3 have been found to have different roles in transcription. In yeast, H3K4 methylation is catalyzed by a single methyltransferase Set1/COMPASS, which performs all states of H3K4 methylation (Ruthenburg et al., 2007). In higher organism, *Drosophila* possesses three and humans six COMPASS family members, each have non-redundant functions (Shilatifard, 2012). H3K4me3 is detected primarily at the TSS to help pre-initiation complex assembly (Vermeulen et al., 2007; Wysocka et al., 2006). In contrast H3K4me1 works in concert with acetylated H3K27 (H3K27ac) to mark active enhancers (Bonn et al., 2012; Creighton et al., 2010; Zentner et al., 2011).

*H3K9me2 and H3K9me3*: Different H3K9 methylation states are mediated by distinct enzymes and are shown to have different functions. Methylation of H3K9 in humans requires the Suv39 family as well as the non-Suv39 enzymes PRDM2 and ASH1L (Volkel and Angrand, 2007). G9a and GLP are H3K9 mono- and di-methyltransferases, while Suv39h is the di- and tri-methyltransferase. Methylated H3K9 is a mark for repressive heterochromatin. It has been shown that mouse Suv39h specifically H3K9me3 at pericentric heterochromatin region (Lachner et al., 2003). Meanwhile in proliferating cells, the repressive signal appears to be primarily G9a-mediated H3-K9 di-methylation (Tachibana et al., 2002).

*H4K20me1/2 and H4K20me3*: H4K20 methylation is a well studied histone mark with separate functions of different methyl states. H4K20me1 is mediated by SET8/PR-Set7 (Fang et al., 2002);

Nishioka et al., 2002), whereas SUV4-20H1 and SUV4-20H2 enzymes produce H4K20me2 and H4K20me3 (Schotta et al., 2004; Schotta et al., 2008). Different H4K20me states have distinct biological functions (Jorgensen et al., 2013): H4K20me1 and H4K20me2 are recognized by 53BP1 to be involved in replication (Tardat et al., 2007; Wu et al., 2010) and DNA damage repair (Botuyan et al., 2006; Sanders et al., 2004) while H4K20me3 marks has been shown to be recognized by L3MBTL1 for heterochromatic region silencing (Gonzalo et al., 2005; Regha et al., 2007; Schotta et al., 2004).

H3K79me2 and H3K79me3: Dot1 in yeast is the sole H3K79 methyltransferase capable of catalyzing all three methyl-states in a non-processive manner (Frederiks et al., 2008). Different H3K79 methylation states also have different functions (Nguyen and Zhang, 2011). For example, in yeast, only H3K79me2 fluctuates with the cell cycle, whereas H3K79me3 levels remain constant (Schulze et al., 2009). Additionally, H3K79me2 and H3K79me3 were found to occur in different regions of the chromosomes in oocytes and enriched in different stage of early fly embryogenesis (Ooga et al., 2008).

H3K36me2 and H3K36me3: Multiple Set2 family members existing in human (Wagner and Carpenter, 2012) suggest the dissected roles of this methylation mark however the detailed mechanism is not known. H3K36me3 is catalyzed exclusively by Set2 ortholog Setd2/Hypb (Edmunds et al., 2008) while NSD family members have been found to confer H3K36 mono- and di- methylation states (Kuo et al., 2011; Li et al., 2009b; Rayasam et al., 2003; Yang et al., 2008). In human, defects of each methylated states cause different phenotype and diseases in human (Wagner and Carpenter, 2012) suggesting they may signal for different pathways. Moreover, H3K36me2 but not H3K36me3 has been shown to be the signaling mark during DNA damage response (Fnu et al., 2011; Jha and Strahl, 2014). In yeast, H3K36me2 has been shown to be less correlated with transcription frequency than H3K36me3 (Rao et al., 2005). Interestingly, H3K36me2 is sufficient for Set2/Rpd3S pathway to repress the initiation of intragenic cryptic promoters (Li et al., 2009a) suggesting H3K36me3 may have non-redundant function.

Here, we aimed to dissect the functions of H3K36me3 and lower methyl-states. By

manipulating the F/Y switch within the catalytic domain of Set2, we generated two mutants that confer different methyl-states at H3K36 both *in vitro* and *in vivo*. We also demonstrated that these two mutants displayed distinct physiological defects in the context of transcription. Lastly, we showed that DNA repair pathway requires different states of K36me as well.

## Results

### **Mutations at the catalytic sites of Set2 alter the specificity of Set2-mediated H3K36 methylation *in vivo***

It has been shown that in yeast H3K36me<sub>2</sub> is sufficient for Set2-Rpd3S pathway suggesting the possible non-redundant function of H3K36me<sub>3</sub> (Li et al., 2009a), however the state-specific function of H3K36me<sub>3</sub> is unclear. To date mutants of other factors have been found to give certain state-specific methylation of H3K36. For instance, deletion of Paf1 complex subunit Paf1 or Cdc73 causes only loss of H3K36me<sub>3</sub> (Chu et al., 2007). However these factors are also found to be involved in multiple pathways (e.g. histone ubiquitylation) therefore it is still hard to pinpoint the exact roles of state-specific H3K36 methylation. We decided to take advantage of F/Y switch mechanism by mutating Set2 catalytic core to control its product specificity. Since “F/Y switch” has been identified in some methyltransferase to direct products specificity (Trievel et al., 2002), we searched for similar module in yeast Set2 (ySet2) by aligning the amino acid sequence of *Saccharomyces cerevisiae* Set2 to those of human Hypb/Setd2 and *Drosophila* Hypb (Bell et al., 2007; Sun et al., 2005). As shown in Fig 4.1 A, yeast Set2 shared significant sequence homology with its higher organism orthologs in both SET domain and the co-factor binding site. Notably, three highly conserved aromatic Y149, F234, and Y236 residues were identified in catalytic center. We modeled the aligned yeast SET domain residues based on human Setd2 SET domain structure (Zheng et al., 2012) (Fig 4.1 B). It shows that Y149, F234, and Y236 are all reside along a narrow channel which is considered as catalytic core. The degraded co-factor OUM localizes at the opposite site of this channel. Y236 sits right next to OUM and is proposed to be co-factor AdoMet and substrate lysine binding residue (Dillon et al.,

2005). To assess the potential roles of these aromatic residues in controlling product methyl-state specificity we decided to mutate Y149 (Me state residue) and F234 (F/Y switch residue). The reason we did not make any alternation on Y236 is that based on the model it coordinates with the co-factor binding and we predicted the mutation of this residue would cause the failure of catalysis.

We first tested the bulk levels of H3K36me<sub>2</sub> and H3K36me<sub>3</sub> in the Y149 and F234 mutants. Using PCR-based mutagenesis, we mutated Y149 to Ala, Asn, Phe and Ser and F234 to Tyr. These mutants were expressed in *SET2* null yeast strains under control of moderately over-expressed *ADH1* promoter. Mutations of Y149 to Phe, Ala or Asn all caused complete loss of H3K36me<sub>2</sub>. Intriguingly, Y149F retained wild type level of H3K36me<sub>3</sub> (Fig 4.2 A) while Y149A and Y149N had comprise H3K36me<sub>3</sub> suggesting that the benzene ring at Y149 site is required for catalysis. Y149S led to loss of both H3K36me<sub>2</sub> and H3K36me<sub>3</sub>, indicating positively charged residue was toxic for the catalysis. On the other hand, F234Y mutant have H3K36me<sub>2</sub> comparable to WT but abrogated H3K36me<sub>3</sub>. Remarkably, our study identified two H3K36me state specific mutants and we defined Y149F mutant as H3K36me<sub>3</sub> only mutant while F234Y as H3K36me<sub>3</sub> defect mutant.

We then tried to investigate whether the auto-inhibitory Set2 middle domain (see Fig 2.5 E) can cross-talk with catalytic core residues to regulate states of H3K36me. To this end, bulk histone H3K36me levels were checked in cells with combined F/Y switch mutations with the AID B region deletion ( $\Delta$ 303-338 a.a.). As shown in Fig 4.2A, combining B region deletion of AID with Y149A or Y149N mutation caused complete loss of H3K36me<sub>2/3</sub>, suggesting disruption of AID may also affected the conformation of catalytic center. This phenomenon was interesting because we showed before deletion of B region of Set2 caused this mutant even more active than WT Set2 (Fig 2.7 C). Moreover, deletion of B region had no effect on F234Y suggesting there is no cross-talk between B region and F234 residue.

We also express these two mutants in WT yeast with the presence of endogenous Set2. At presence of WT Set2, expression of Y149F still caused total loss of H3K36me<sub>2</sub> while F234Y

significantly reduced H3K36me3. This may be due to that moderate over-expression of *ADHI* promoter driven Set2 mutants competed with the endogenous WT Set2 (Mumberg et al., 1995). For instance, Y149F, H3K36me3 only mutant, would convert all naturally existing H3K36me1/2 into H3K36me3. On the other hand, F234Y, the H3K36me3 defect mutant, would compete with endogenous WT Set2 for elongating Pol II occupancy and abrogate H3K36me3.

**Set2 catalytic pocket mutants show intrinsic state-specificities of H3K36 methylation *in vitro*.**

We next want to check the biochemical properties of these mutant Set2. Recombinant Y149F and F234Y Set2 mutants were purified from insect cells and subjected to HMT assays using reconstituted nucleosomes. The results showed that Y149A, Y149N, and Y149F mutants produced only H3K36me3, whereas F234Y yielded the minimal H3K36me3 and normal level of H3K36me2 and H3K36me1 (Fig 4.3 A). This result is consistent with the bulk levels of H3K36me detected *in vivo*. Although the steady state levels of H3K36me3 and H3K36me2 produced by Y149F or F234Y, respectively, were similar to WT levels (Fig 4.2 A), we were interested in whether the enzyme kinetics of these two mutants were altered. To this end, we plot the enzymatic activity to the concentration of Set2 in HMT assays. H3K36 state-specific antibodies were used to measure the progression of each methyl state of H3K36. As result shown, the mutant enzymes have different enzymatic kinetics compared to WT (Fig 4.3 B). WT Set2 shows both features of processive and distributive enzyme. Because from the lowest concentration, almost all the products directly end up with H3K36me2 but not H3K36me1 indicating this enzyme stays on the substrate until completion of di-methylation (processivity). Then with the increased WT Set2 concentration, H3K36me3 and H3K36me2 are both accumulated suggesting under these conditions the enzyme displays the feature of distributive enzyme in which the co-factor exchange concurs with the dissociation of enzyme from the substrate (Frederiks et al., 2008). Meanwhile F234Y acts solely like a distributive enzyme which produces H3K36me1 and H3K36me2 proportionally as concentration increases. By contrast, Y234F mutant displayed the

characteristics of a pure processive enzyme because only H3K36me<sub>3</sub> was accumulated. Moreover, the HMT assay using <sup>3</sup>H labeled co-factor SAM shows the total methyl-group incorporation rates are also different between WT and mutant enzymes (Fig 4.3C). F234Y showed significantly elevated reaction rate than WT. However, Y149F has much lower reaction rate compared with WT. In summary, our results suggest that aromatic residues Y149 and F234 in the catalytic core module of SET domain largely control the accommodation of methylated lysine during catalysis. Besides altering the product specificities, the point mutation of F149 or Y234 also results in significant changes of enzyme kinetics.

**Set2 Y149F (me<sub>3</sub> only) and F234Y (me<sub>3</sub> deficient) mutants do not alter their p-Pol II binding properties.**

To investigate whether the changes of H3K36me in Y149F or F234Y mutants were due to abnormal recruitment of Set2 to p-CTD, we performed *in vitro* p-CTD peptide pull-down assays of both mutants. Set2/p-CTD interaction has been shown to correlate the Set2 stability and H3K36 methylation levels in cells (Fuchs et al., 2012). However, this was not likely the case for Y149F or F234Y mutants because the di- or tri- methylation levels they produced, respectively, were comparable to that of WT. Indeed, we showed that recombinant Y149F or F234Y had comparable p-CTD binding abilities to WT Set2 (Fig 4.4). Therefore, unlike SRI domain deletion (Chu et al., 2007; Fuchs et al., 2012), our H3K36me state specific mutants may have normal H3K36me distribution patterns because they still interact with elongating Pol II.

**Set2 catalytic pocket mutants show cryptic transcription phenotype.**

Previous discovery showed that H3K36<sup>2/3</sup> recruited Rpd3S deacetylation complex onto the transcribed genes to suppress the initiation of cryptic promoters (Carrozza et al., 2005b; Li et al., 2007b). And it was also found H3K36me<sub>2</sub> is sufficient for this pathway (Li et al., 2009a). A handful of studies also tried to dissect the contribution of individual H3K36 methylation state in

this pathway. However, those studies mainly focus H3K36me<sub>2</sub> due to the lack of mutant with only H3K36me<sub>3</sub>. To check the contribution of individual H3K36 methylation state in Set2/Rpd3S pathway, we checked the cryptic transcription phenotype on the *STE11* gene in Y149F or F234Y mutants using northern blotting. Surprisingly, we observed that the catalytic core mutants all displayed cryptic transcription phenotype at 3' end of *STE11* (Fig 4.5 A). To confirm this result, we also tested this phenotype in a more sensitive reporter system. In this *STE11-HIS3* reporter system, *HIS3* gene was integrated at the downstream of the *STE11* cryptic promoter. The *HIS3* coding sequence was inserted out-of-frame with respect to the *STE11* coding sequence. The functional *HIS3* mRNA can only be made when transcription is initiated at the cryptic start site where the ATG of *HIS3* will be used for translation. This will result in viable cells on synthetic media lacking histidine (Fig 4.5 B). Plasmid-borne Set2 WT, Y149F, and F234Y were expressed in either WT or  $\Delta SET2$  background. Consistent with Northern blotting, both Y149F and F234Y fail to suppress the cryptic transcription of *STE11* in  $\Delta SET2$  background and caused growth on SD-His plates (Fig 4.5 C). Notably, the H3K36me<sub>3</sub> defective F234Y had a more severe phenotype than H3K36me<sub>3</sub> only Y149F mutant (Fig 4.5 C). In the presence of WT Set2, phenotype of F234Y was alleviated while that of Y149F completely rescued (Fig 4.5 C). In conclusion, using two independent systems, we showed that both mutants are not sufficient to suppress cryptic transcription at *STE11*.

### **Set2 state-specific mutants display different synthetic phenotypes with histone chaperone yFACT mutant.**

To test if these two mutants display other distinct functions, we looked into the known genetic interactions of Set2 with other transcription related mutants. It has been shown that the growth defect of yeast histone chaperone yFACT subunit mutant *spt16-11* can be rescued by deletion of *SET2* (Biswas et al., 2006). Therefore, we tested whether Y149F or F234Y behave differently in histone chaperone pathway using ectopic expressed mutants. Consistent with previous observation,  $\Delta SET2$  rescued the growth defect of *spt16-11* at semi-permissive

temperature (35°C) while FL Set2 made *spt16-11* sicker under same condition (Fig4.6 A). H3K36me3 defective F234Y mutant functioned similarly to FL Set2, meaning H3K36me1/2 was enough for maintain Set2's role in yFACT pathway (Fig4.6 A). To our surprise, H3K36me3 only Y149F mutant suppressed the growth of *spt16-11* even stronger than FL Set2. This may indicate that the dynamic and balance between H3K36me1/2 and H3K36me3 are important for this pathway.

We then examined whether the defects caused by Y149F or F234Y depend on the Rpd3S pathway. To this end, we deleted *RCO1*, an essential gene in Set2/Rpd3S pathway (Carrozza et al., 2005a), in the above mutant background. Previous studies show that deletion of *RCO1* rescue the growth defect of *spt16-11* (Biswas et al., 2008a). Our growth assays showed that deletion of *RCO1* rescued the synergistic growth defects between *spt16-11* and all Set2 constructs tested (Fig 4.6 B). The fact that deletion of *RCO1* can bypass H3K36me3 only Y149F mutant suggested that Y149F caused gain-of-function defects that work through the Rpd3S pathway.

### **Set2 state-specific mutants display different synthetic phenotypes with prolyl-isomerase Ess1 mutant.**

We next tested another known genetic interaction of Set2. Ess1 is one of Peptidyl-prolyl isomerases (PPIases) in yeast which targets RNA Pol II CTD phospho-Ser5 to control Ser5-Pro6 bond *cis-trans* peptide isomerization (Hanes et al., 1989; Hani et al., 1995). Previous study shows that deletion of *SET2* rescues the growth defect of *ess1<sup>H164R</sup>* mutant at non-permissive temperature (Ma et al., 2012), suggesting genetic interaction between them. To test whether Y149F or F234Y has synthetic effect with Pol II prolyl-isomerase Ess1, growth assays under semi-permissive temperature were perform in WT, Y149F or F234Y mutants with *ess1<sup>H164R</sup>/ΔSET2* background. Consistent with published data, deletion of *SET2* significantly rescued the growth defect of *ess1<sup>H164R</sup>* while expressing WT Set2 suppressed the growth (Fig3.7). Notably, Y149F (H3K36me3 only) also suppressed the growth like WT Set2. In contrast, F234Y (H3K36me3 defect) behaved like empty vector. Therefore, similar with the studies on yFACT

pathway, distinct H3K36me states seem to cause different phenotype. However, Y149F which confers only H3K36me<sub>3</sub> acts like WT while F234Y which defects on H3K36me<sub>3</sub> abolishes Set2 function in this pathway.

### **DNA Double-strand break repair requires specific states of H3K36me.**

Recent studies in yeast suggest that H3K36 methylation also plays roles in checkpoint activation and DNA double-strand break repair (Jha and Strahl, 2014; Pai et al., 2014). Therefore we wanted to test if these two mutants cause differential defects in DNA double-strand break repair. We used plasmid re-ligation assay to measure the repair efficiencies of linearized pRS415 plasmids by HindIII digestion of WT and mutant Set2 (Fig 4.8 A). In our system, linearized pRS415 plasmids were transformed into cells with either WT or Y149F or F234Y mutant cells. Repair efficiency was calculated by the rate of transformants produced by either linearized or circular plasmid (Fig 4.8 C). Deletion of *SET2* increased the repair rate compared to WT (Fig 4.8 D), which is consistent with published data (Fnu et al., 2011). Interestingly, H3K36me<sub>3</sub> defect mutant F234Y showed wild type level of repair efficiency while H3K36me<sub>3</sub> only mutant Y149F displayed elevated plasmid repair rate similar to *SET2* deletion. Our data suggests that cells utilize H3K36me<sub>1/2</sub> for NHEJ repair and excess H3K36me<sub>3</sub> may interfere with this pathway.

## Discussion

In this study, we discovered the conserved F/Y switch residues in yeast Set2 SET domain by structural prediction based on the sequence homology and structure modeling on known human Setd2/Hypb SET domain. Taking away hydroxyl group from conserved Tyr in catalytic core enlarges the diameter of the active site, accommodating the increasing size of the methylated  $\epsilon$ -amino group during successive methyl transfer reactions (Couture et al., 2008; Del Rizzo et al., 2010). Thus the Tyrosine hydroxyl group is considered to generate steric hindrance to restrict multi-round methyl-transferring catalysis (Collins et al., 2005). And this mechanism is conserved in yeast Set2 to control product specificity.

We found that Y149F mutant displays only tri-methyltransferase activity with highly compromised reaction rate (Fig 4.3 B, C) suggesting H3K36me3 is the rate limiting step of the entire methylation reaction. And this feature may be important for the timing issue of H3K36me3 related signaling pathways. Similar result was seen by other group showing in SET7/9 Y245A convert the WT enzyme to a tri-methyltransferase with weak mono-methyltransferase activity (Xiao et al., 2003a). The later study showed that the turnover number of this mutant for un-methylated peptide substrate was diminished over 10 folds *versus* methylated substrates and reduced over 30 folds *versus* the WT enzyme. This suggests the potential mechanism of reaction rate reduction in tri-methyltransferase mutant (Del Rizzo et al., 2010).

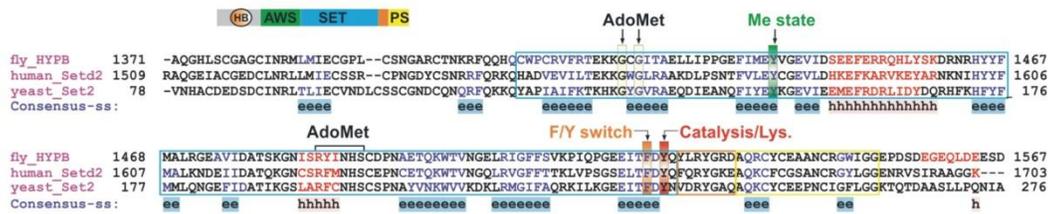
By using Set2 catalytic core mutants which bear minimal alternation on enzyme and no genetic background change, we have shown that either H3K36me3 or H3K36 me2 alone is not sufficient to suppress cryptic transcription at *STE11* in two independent systems. Our *STE11-HIS3* reporter system further suggests that the H3K36me3 and H3K36me2 cooperatively function in Set2-Rpd3S pathway and H3K36me3 contributes more than H3K36me2.

Utilizing these methyl-state specific mutants, distinct phenotypes between Y149F and F234Y were observed in several pathways including histone chaperone (Figure 4.6), CTD proline isomerization (Figure 4.7), and double-strand DNA damage repair (Figure 4.8). Due to the complexity of these pathways, we speculate: 1. Altered enzyme kinetics may be involved in the

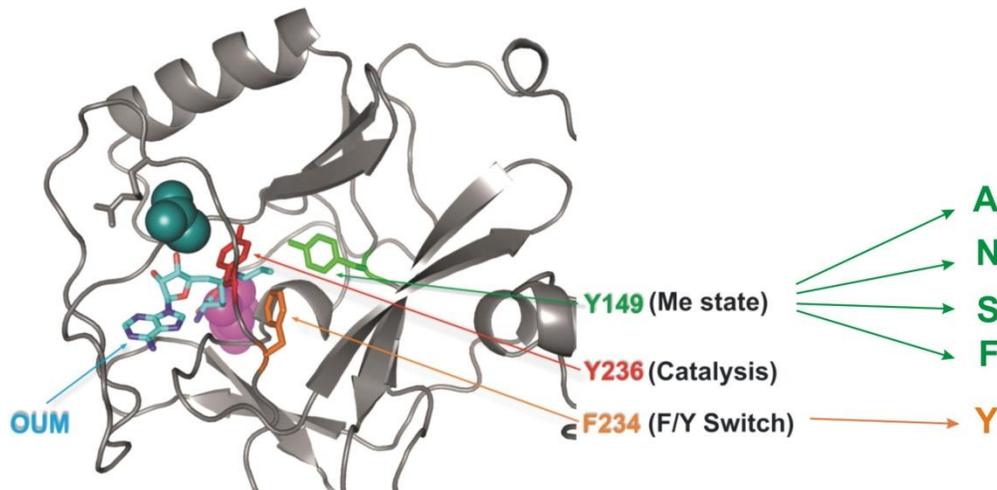
phenotypes of cryptic transcription suppression in Y149F and F234Y mutants (Fig4.3 C). 2. Different methylation states may affect histone exchange which affected the cryptic transcription repression (Smolle et al., 2012; Venkatesh et al., 2012). It also helps explain the synthetic growth phenotypes of Y149F and F234Y mutants combined with histone chaperone Spt16. 3. The double-strand DNA damage assay indicated that H3K36me2 but not H3K36me3 is the mark to signal for repair which is in agreement with what found in both yeast and human (Fnu et al., 2011; Jha and Strahl, 2014).

The signal of histone modifications are recognized by specific downstream readers (Yun et al., 2011). And we believed that state-specific methylation may be read by different downstream factors. It has been shown that chromo and PWWP domains containing proteins recognize H3K36 methylation (Yun et al., 2011). For example, DNA damage repair factor 53BP1 and its homolog in fission yeast Crb2 specifically recognize H4K20me2 by tandem tudor domains but do not bind H4K20me3 (Botuyan et al., 2006). Among H3K36 methylation binding factors, DNMT3A and MSH-6 are shown to specifically recognize H3K36me3 rather than other H3K36 methylation marks, rendering possibility that these factors can potentially mediate state-specific function of methylated H3K36 (Dhayalan et al., 2010; Li et al., 2013). We believe that more and more factors that have methyl-state preferences binding for H3K36 or other histone methylation marks will be discovered and studying their function may help us understand how histone codes are recognized by readers to reach distinct functional outputs.

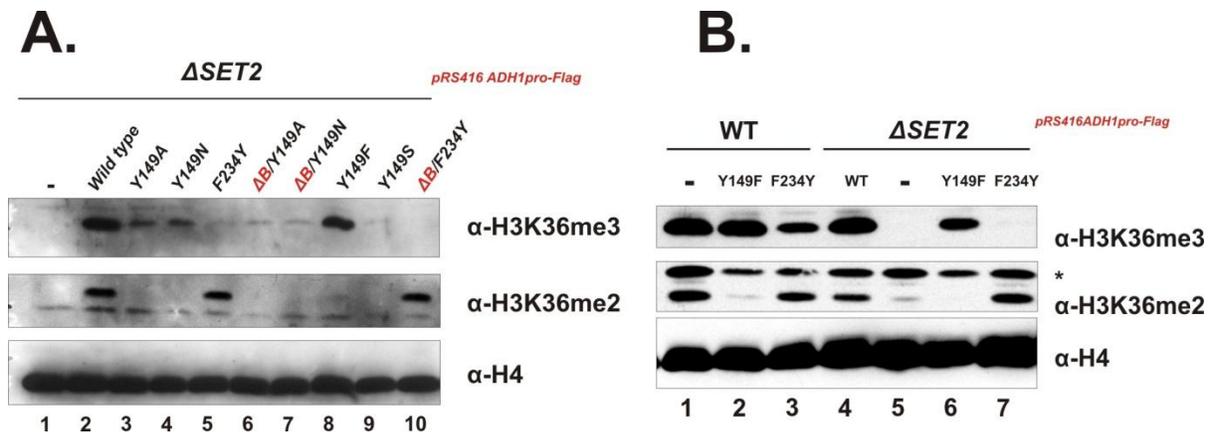
A.



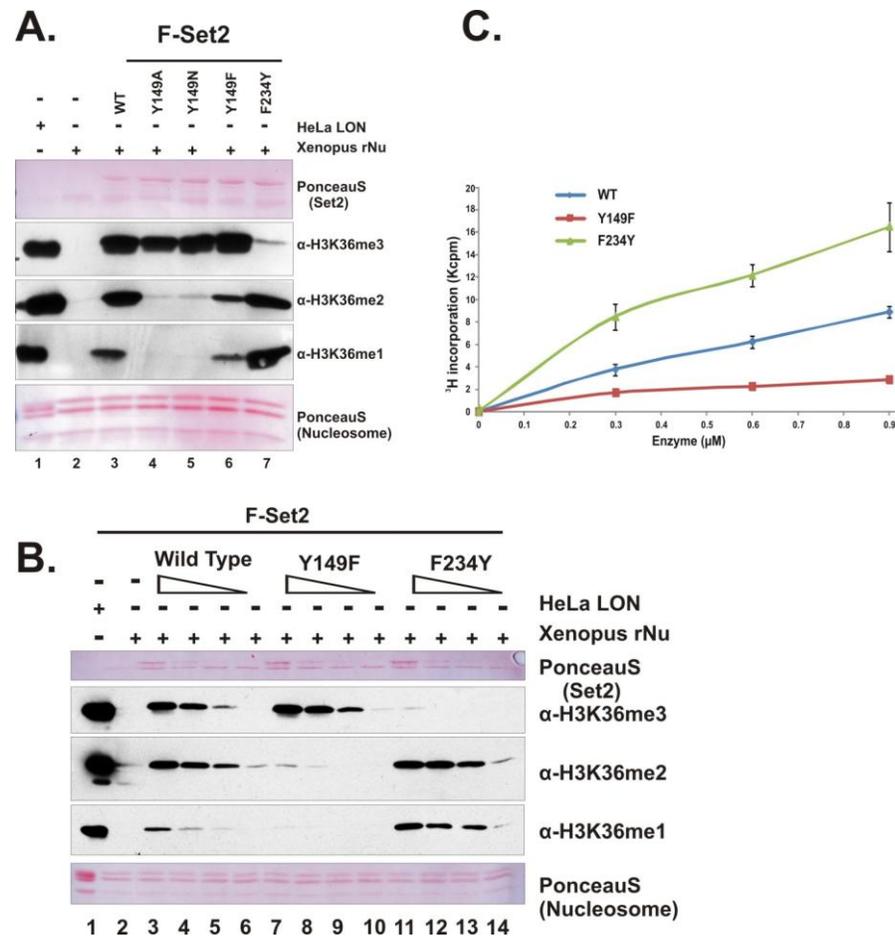
B.



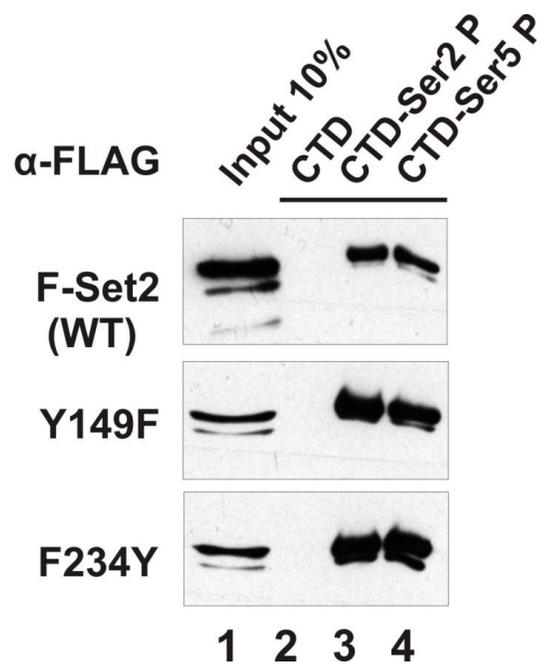
**Figure 4.1** *Set2* contains conserved residues that are predicted to control state-specificity. (A) Sequence alignment of the SET domains of *Drosophila melanogaster* Hypb, human Hypb/Setd2, and *Saccharomyces cerevisiae* Set2. The residues of SET domain, post SET loop and post SET domain are marked in blue, orange, and yellow boxes respectively. The co-factor (AdoMet) binding sites are indicated by black arrows and lines. Key residues within catalytic pocket which control methylation state (Me state), Phe/Tyr switch (F/Y switch) and catalysis (Catalysis/Lys) were indicated in green, orange, and red respectively. The residues that were predicted as consensus helix motif were highlighted in red and indicated using letter “h” while residues predicted as extended strands were highlighted in blue using letter “e”. (B) The structure model of ySet2 SET domain were constructed based on human SetD2 SET domain (PDB 4FMU) using Swiss-modelling. The three key aromatic residues in the catalytic pocket were highlighted also in red, yellow and green as shown in (A). The degraded product of co-factor SAM (OUM) and two zinc atoms in the crystal were shown as colorful sticks, green or purple spheres, respectively.



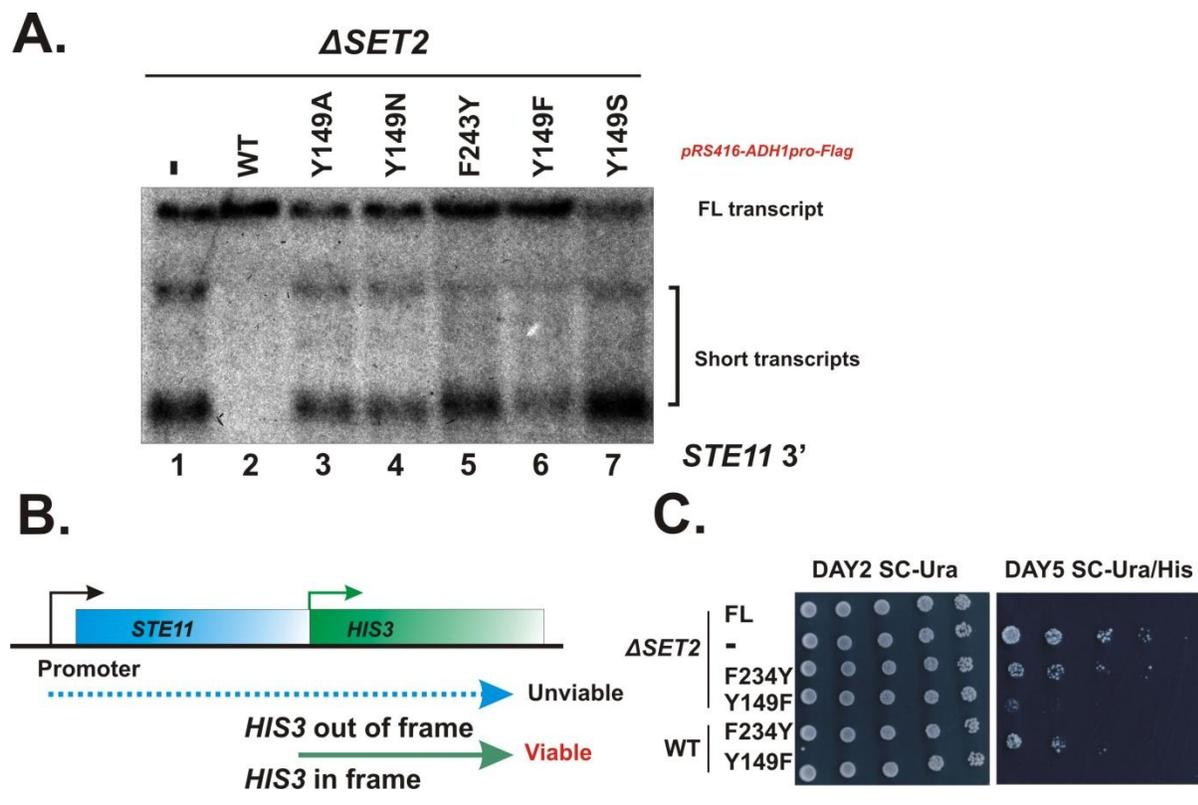
**Figure 4.2 Mutations at the catalytic sites of Set2 alter Set2-mediated H3K36 methylation in vivo.** (A) Western blotting analysis of H3K36 methylation states on bulk histones from cells expressing Set2 catalytic pocket mutants. Indicated antibodies were applied to detect specific methylation states. Plasmids carrying only FLAG-adaptor sequence (pWY084), the ORFs of wild type *SET2* (pWY088) or variants: Y149A (pWY157), Y149N (pWY158), F234Y (pWY159), Y149S (pWY164), Y149F (pWY165),  $\Delta B$  ( $\Delta 303$ -338)/Y149A (pWY161),  $\Delta B$  ( $\Delta 303$ -338)/Y149N (pWY162),  $\Delta B$  ( $\Delta 303$ -338)/F234Y (pWY163) were transformed into *SET2* null yeast strain (YYW010). All of the yeast Set2 expressing vectors contain pRS416 backbones inserted with an oligo adaptor containing *ADH1* promoter, 1x *FLAG* tag sequence at 5'end, and *CYC1* terminator at 3'end (pRS416 ADH1pro-Flag (pWY084)). Point mutations on catalytic pocket residues alone or combined with the deletion of B region in the Set2 middle domain (aa303-338) were introduced into the *SET2* ORF. (B) Western blotting analysis showing Set2 Y149F and F234Y mutants have distinct H3K36 methylation states, and over-expressed mutant proteins functionally compete with endogenous WT Set2. The asterisk indicates the cross-reaction of anti-H3K36me2 antibody.



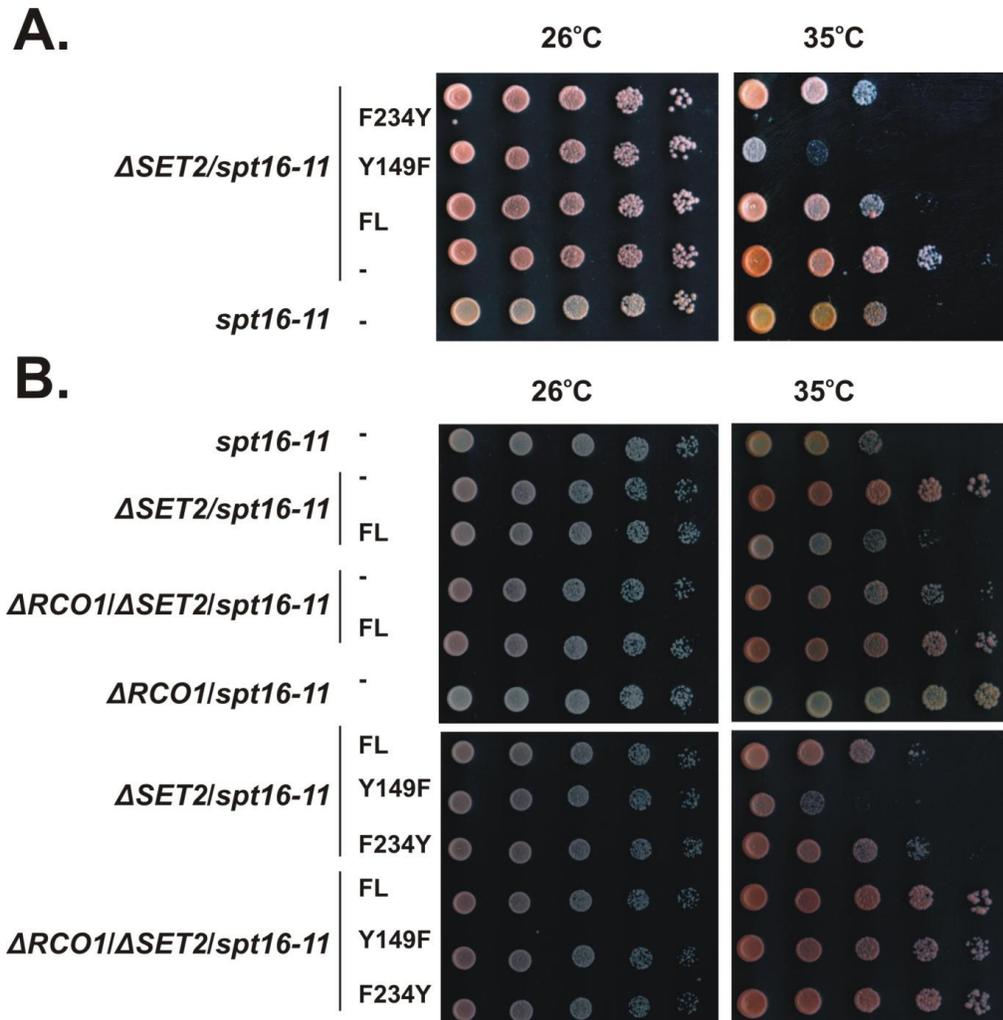
**Figure 4.3** *Set2* catalytic pocket mutants show intrinsic state-specificities of H3K36 methylation *in vitro*. (A) Western blotting analysis showing intrinsic H3K36 methylation state-specificities of Set2 catalytic pocket mutants on reconstituted nucleosomes. The *In vitro* Histone Methyltransferase Assays (HMT) were performed using equal amounts of F-Set2 variants purified from insect cells, recombinant *Xenopus* nucleosomes, and co-factor S-Adenosyl methionine (SAM). The indicated antibodies were applied to detect individual H3K36 methylation states. Ponceau S staining shows the loading of F-Set2 proteins and nucleosome substrates. The HeLa LON nucleosomes were used as positive control for Western blotting assay. (B) Western blotting analysis showing altered enzyme dynamics of Y149F and F234Y mutants at each individual state of H3K36 methylation. HMT assay were performed using equal molar of F-Set2 variants recombinant, *Xenopus* nucleosomes and co-factor S-Adenosyl methionine (SAM). The reactions were subjected to Western blotting analysis with the indicated antibodies. Ponceau S staining shows the loading of F-Set2 proteins and nucleosome substrates. The HeLa LON nucleosomes were used as positive control for Western blotting. (C) Standard HMT assays using  $^3$ H labeled SAM to show the different  $^3$ H incorporation rates between wild type Set2 and mutants.



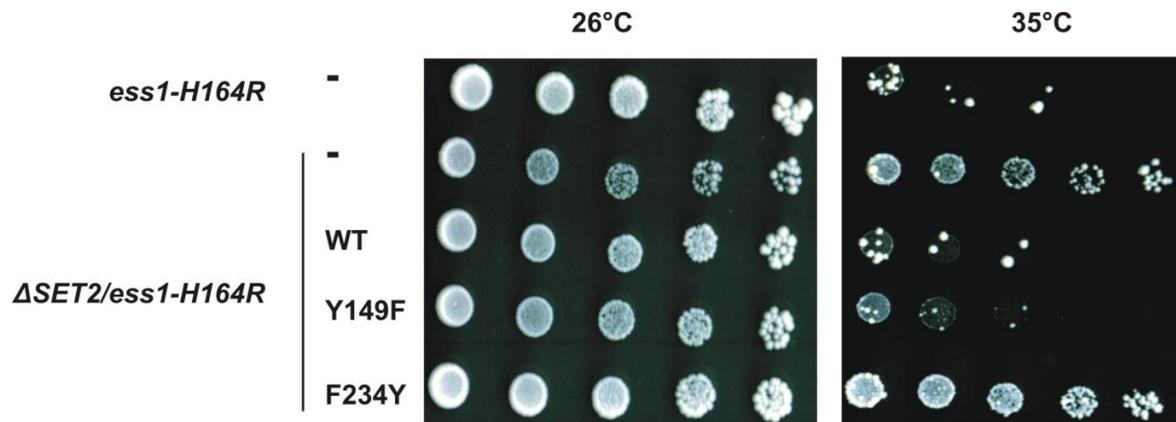
**Figure 4.4** *Set2 Y149F (H3K36me3 only) and F234Y (H3K36me3 deficient) mutants do not alter their p-Pol II binding properties.* CTD peptide pull-down experiments showing binding between Set2 variants and Pol II CTD peptide. *In vitro* pull-down assay was performed using recombinant FLAG-Set2 variants and biotinylated CTD peptides: unphosphorylated (CTD), phosphorylated on Ser5 (Ser5P) or Ser2 (Ser2P). Streptavidin-coated magnetic beads were used to pull down CTD peptide bound Set2 variants and Western blot analysis was performed using anti-FLAG antibody.



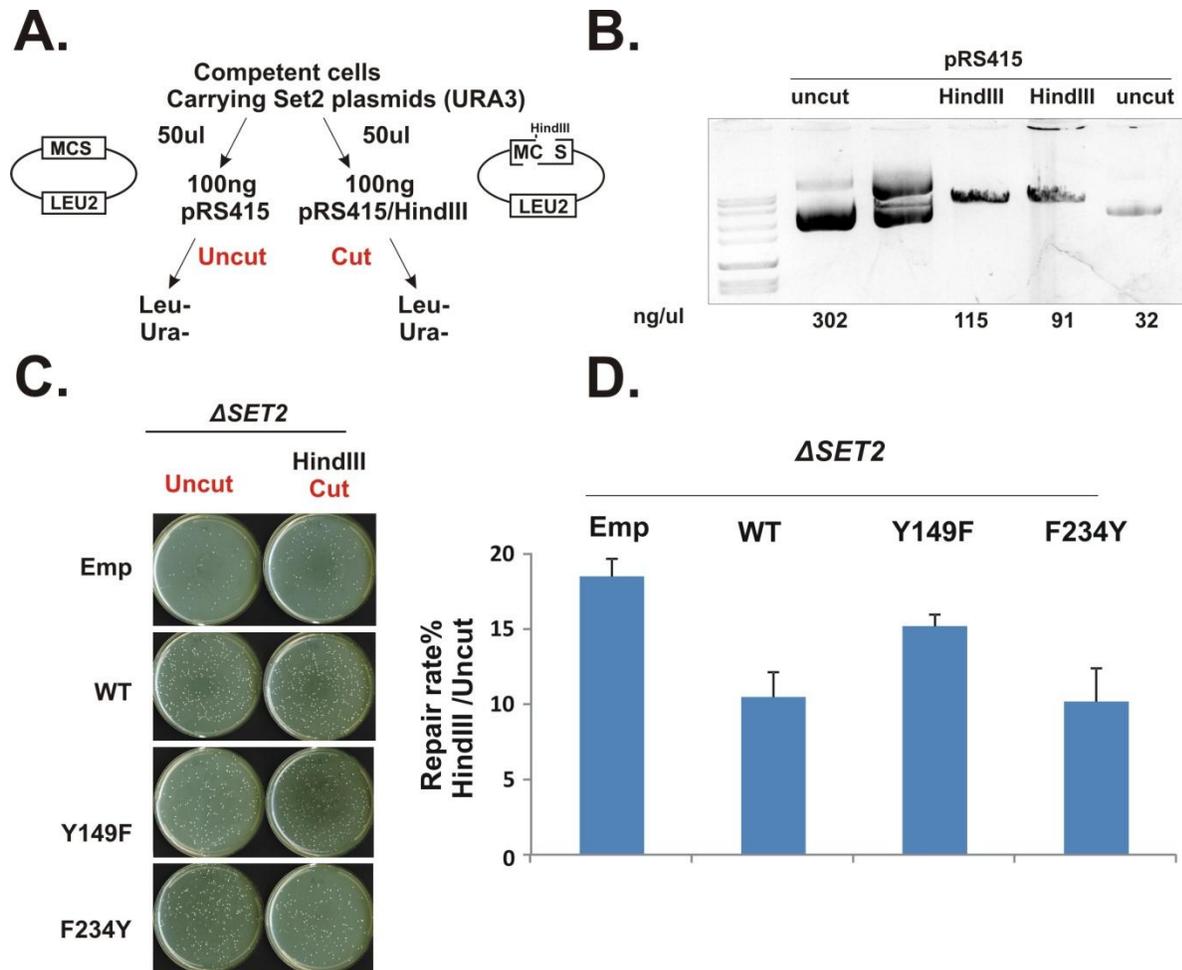
**Figure 4.5** *Set2* catalytic pocket mutants cause intragenic spurious transcription of *STE11*. (A) The Northern blot analysis of *STE11* transcripts from *SET2* null yeasts expressing plasmid-borne WT *Set2* or catalytic pocket mutants using *STE11*-3'RNA probe labeled through *in vitro* transcription. Full length and two short transcripts of *STE11* are indicated. (B) A diagram of cryptic transcription reporter strains in which *HIS3* genes were integrated downstream of the *STE11* cryptic promoter (*STE11-HIS3*). The integration sites were selected such that the *HIS3* gene is out of frame with relation to the *STE11* coding region, and the functional His3 can only be produced when the *HIS3* transcript initiates at the cryptic promoter of *STE11* (Ruan and Li, unpublished). This strategy was first developed by the Winston lab (Cheung et al., 2008). (C) The spotting assay showing the initiation of *STE11* cryptic promoter by the growth on SC-Ura/SC-His plates. Indicated plasmid-borne WT *Set2* or mutants were transformed into the reporter strains YCR376 (*STE11-HIS3*) and YCR377 (*STE11-HIS3*/ $\Delta SET2$ ). The resulting strains were grown on SC-Ura or SC-Ura/SC-His plates at 30°C for 2-5 days.



**Figure 4.6** *Set2* state-specific mutants display different synthetic phenotypes with histone chaperone *yFACT* mutant. (A) *Set2* Y149F causes synthetic growth phenotype with *spt16-11* mutant at 35°C (semi-permissive temperature) while F234Y behaves like WT *Set2*. Either WT or catalytic pocket mutant plasmids were transformed into  $\Delta SET2/spt16-11$  yeast strain (YYW163). Transformants were grown at 26°C until saturation. 5-fold serial dilutions of each transformants (from OD 600=1.0) were then spotted on to SC-Ura plates and grown at 26°C (permissive temperature) and 35°C, respectively. Photographs of the plates were taken after 2-3 days. (B) Deletion of *RCO1* rescues the synthetic growth defect between *Set2* Y149F mutant and *Spt16-11* by bypassing the requirement of *Set2*. WT and catalytic pocket mutant plasmids were transformed into  $\Delta SET2/spt16-11$  (YYW163) or  $\Delta RCO1/\Delta SET2/\Delta RCO1/spt16-11$  (YYW165) yeast strains respectively. Transformants were grown at 26°C until saturation. 5-fold serial dilutions of each transformants were then spotted onto SC-Ura plates and grown at 26°C and 35°C. Photographs of the plates were taken after 2-3 days.



**Figure 4.7** *Set2* state-specific mutants display different synthetic phenotypes with RNA Pol II prolyl-isomerase *Ess1* mutant. WT or catalytic pocket mutant plasmids were transformed into *ess1-H64R/ΔSET2* (YYW160) yeast strain. Transformants were grown at 26°C until saturation. 5-fold serial dilutions of each transformants were then spotted onto SC-Ura plates and grown at 26°C and 35°C. Photographs of the plates were taken after 2-3 days.



**Figure 4.8 DNA Double-strand break repair requires specific states of H3K36me.**

(A) An experimental scheme of plasmid repairing assay to measure the repair efficiency of HindIII linearized pRS415 plasmid (*LEU2*) in  $\Delta SET2$  yeasts (YYW 010) expressing different plasmid-borne Set2 mutants (*URA3*). Competent cells were transformed with 100ng of pRS415 plasmid with/without HindIII digestion and spread onto uracil/leucine synthetic double drop-out plates. The colony numbers of both circular and linearized pRS415 transformants were counted. (B) Agarose gel showing the completion HindIII digestion and quantification of pRS415 plasmid. (C) The transformants with uncut or HindIII cut pRS415 plasmid growing on SC-Ura/SC-Leu double selection plates for 2 days. (D) The plasmid repair rates of each Set2 variant strains. The plasmid repair rate is represented by the percentage ratio of colony numbers between HindIII-cut and un-cut transformants.

## Materials and Methods

### Construction of Plasmids and Yeast Strains:

The yeast expression plasmid harboring *ADHI* promoter and *CYCI* terminator sequences (pWY080, a gift from Dr. Brian Strahl) was modified by inserting a short DNA adaptor containing XbaI-ATG-FLAG-XhoI-SmaI-NotI-Sall sequences to form pRS416 ADH1pro-Flag vector (pWY084). The *SET2* ORF was generated by the standard PCR reaction from genomic DNA. The resulting ORF fragments containing XhoI cutting site at 5' and NotI site at 3' site were cloned into pWY084 vector. *SET2* point mutant plasmids were generated by PCR based Site-directed mutagenesis using QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). The PCR reaction for each mutagenesis was performed in a 50µl system containing 100 ng of WT Set2 (pBP-HFT-Set2, pWY088) or  $\Delta B$  mutant (pBP-HFT-Set2  $\Delta B$  ( $\Delta 303-338$ ), pWY105) plasmid DNA template, 5µl of 10x reaction buffer, 1µl of 10mM dNTP, 0.5µl of 20mM primers, and 1µl of PfuTurbo DNA polymerase (Stratagene). Each of 18 reaction cycles went through 95 °C for 30 seconds, 55 °C for 1 minute, and the 500bp/min extension (in this case 16 minutes) at 68 °C. 10µl of total reaction was loaded onto 1% Agarose gel to check the PCR product and 25µl was taken for 1µl DpnI digestion (NEB) at 37 °C for 1 hour. Half of DpnI-treated products were transformed into 50µl of TOP 10 *E coli*. competent cells (Life technologies) by using standard transformation protocol. Each of the mutated plasmids was extracted by using Quick Plasmid Miniprep Kit (Invitrogen), and sequencing confirmed. To construct the FLAG-tagged WT Set2 plasmid for the baculovirus system, the *SET2* ORF was generated by standard PCR reaction from genomic DNA. The 5' primer contains XhoI cutting site while the 3' primer contains NotI site. The fragment was first sub-cloned into pCR™-Blunt vector (Life technologies) and then cloned into XhoI/NotI digested pBacPAK-HFT vector (pBacPAK-8-N-HisFlag-TEVsite, pBL532). For Set2 mutant plasmids, all the fragments were cut from yeast expression vectors by XhoI/ NotI and directly ligated into pBacPAK-HFT vector.

To make *STE11-HIS3* reporter strain (YCR376), *STE11* gene of BY4741 yeast was truncated at specific sites at 3', which was known from the previous publication (Li et al. 2007). The

truncation was made by the replacement of *STE11* 1840-2153 with an I-SceI CORE cassette carrying *URA3* and *G418* marker (Storici F et al. 2001). The integrated CORE cassette was then popped-out by *HIS3* gene fragment flanked by *STE11* homologous sequences using 5-FOA counter selection strategy. To knock-out *SET2* from *STE11-HIS3* reporter strain (YCR377), the gene replacement cassette carrying hygromycin B selection marker flanked by *SET2* homologous sequences was obtained by PCR reaction and transformed to replace *SET2* gene.

The *spt16-11*(YYW162) and *spt16-11/ΔSET2* (YYW163) yeast strains were gifts from Dr. David Stillman and the *ess1-H164R* (YYW 157) was requested from Dr. Steven D Hanes. To make *spt16-11/ΔSET2/ΔRCO1* (YYW165), *spt16-11/ΔRCO1* (YYW164), and *ess1/ΔSET2* (YYW160) yeast strains, the gene replacement cassettes for *RCO1* or *SET2* were generated respectively by PCR amplification. The standard yeast transformation protocol was applied. Briefly, 2μg of purified DNA fragments were mixed with 50ul of yeast competent cell solution, 5ul of sonicated salmon sperm DNA (10ug/ul invitrogen), and 300 ul of yeast plate solution(0.1M LiOAc,10mM Tris HCl pH 7.5, 1mM EDTA pH 8.0 and 40% PEG 3500). After 30 minutes of incubation at RT and 15 minutes heat shock at 42 ° C, cells were spread onto selection media plates. For the antibiotic selection cassette (Hygromycin), heat shocked yeasts were re-suspended in 1ml YPAD liquid culture and incubated at RT for 6 hours before spreading. Colony PCR was performed to screen the positive colonies using specific primer pairs.

### **Preparation of whole-cell extracts for immunoblotting:**

Cells were grown in 3ml synthetic URA drop-out media supplemented with 2% dextrose or 2% Galactose at 30°C for overnight incubation until OD600 was reached 1-1.2. Cell pellets were re-suspended in 45μl of STE buffer (containing 500mM NaCl, 10mM Tris HCl pH8.0 and 1mM EDTA) and 40μl of 3xSDS buffer. 100μl of 0.5mm glass beads were then added and boiled at 95°C for 5 minutes. Samples were then vigorously vortexed and cell suspensions were clarified by spinning at 14000rpm for 5 minutes. Anti-H3K36me3 (Abcam, 9050), anti-H3K36me2 (Abcam, 9049), anti-H4 (Abcam, 10158), and anti-FLAG HRP (Sigma, A8592) antibodies were

used according to manufacturers' suggestions. For anti-H3K36me2 antibody, 5% BSA was used for blockage, and 1% BSA was present in all antibody incubation.

### **Histone Methyltransferase Assays (HMT):**

Standard HMT reactions were carried out in a 40 $\mu$ l system containing 1xHMT buffer (50 mM Tris.HCl pH 8.0, 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM DTT, 5 % Glycerol), 4 $\mu$ g or 3-fold serial dilution of recombinant Set2 proteins, 4 $\mu$ g of Xenopus recombinant nucleosomes and 10 $\mu$ M cold S-Adenosyl methionine (Sigma) (Li et al., 2003b). The reactions were performed at 30°C for 1 hour. To detect specific methylation states on given histone substrates, Western blotting analysis was performed using antibodies against H3K36me3 (Abcam 9050), H3K36me2 (Abcam 9049), and H3K36me1 (Abcam 9048).

For the HMT assay to measure enzyme kinetics, reactions were carried out in a 20 $\mu$ l system. 0.5-2 $\mu$ g of recombination Set2 variants were incubated with 600ng of Xenopus recombinant nucleosomes and 0.5 $\mu$ l of radio-labeled <sup>3</sup>H S-Adenosyl methionine (80 Ci/mmol, Perkin Elmer) at 30°C for one hour. Scintillation counting-based filter-binding assay was performed as described previously (Eberharter et al., 1998).

### **Northern Blot Assay:**

Total RNA was extracted by vortexing cells in 200 $\mu$ l of 0.5mm glass beads(Biospec), 320 $\mu$ l of RNA Prep Buffer (0.5M NaCl, 0.1M Tris-HCl pH 7.5, 10mM EDTA and 1% SDS), and 250 $\mu$ l of Phenol-Chloroform Isoamyle-alcohol (Fisher). RNAs were then Ethanol-precipitated, resolved in 1% agarose-formaldehyde gels, and transferred to Zeta-probe membrane (Bio-Rad). UV-crossed membranes were subjected to hybridization in Denhardt's based buffer (6x SSC, 5X Denhardt's solution, 0.5% SDS, and 0.1 mg/ml of sonicated salmon sperm DNA). 3' *STE11* probe was prepared through *in vitro* transcription reaction using T7 RNA polymerase (MEGAscript® Kit, Life Technologies Corporation) and 5 $\mu$ l of  $\alpha$ -<sup>32</sup>P labeled CTP (Perkin Elmer).

**Spotting Assay:**

The *spt16-11*(YYW162), *spt16-11/ΔSET2* (YYW163), *spt16-11/ΔRCO1* (YYW164), *spt16-11/ΔRCO1/ΔSET2* (YYW165), *ess1-H164R* (YYW157), and *ess1-H164R/ΔSET2* (YYW160) yeast strains harboring pRS416 ADH1pro-Flag empty vector (pWY084), *SET2* WT (pWY088), Y149F (pWY165), or F234Y (pWY159) were cultured in uracil synthetic drop-out medium at 26°C and collected once saturated. The cells were diluted to OD<sub>600</sub>=1.0 followed by further 5-fold serial dilutions. 3ul of diluted cells were spotted onto uracil synthetic drop-out plates to grow at 26°C or 35 °C. In the assay to measure cryptic transcription of *STE11-HIS* reporter gene, *STE11-HIS3* (YCR376) and *STE11-HIS3/ΔSET2* (YCR377) reporter strains transformed with same series of Set2 expression vectors were cultured at 30°C in uracil synthetic drop-out medium overnight. The saturated cell culture was diluted and spotted onto uracil or uracil/histidine double synthetic drop-out plates (SC-Ura/SC-His). Photographs of the plates were taken after 5 days.

**Plasmid re-ligation Assay:**

Competent cells were made using *ΔSET2* yeast (YYW005) expressing plasmid-borne Set2 variants (pRS416 ADH1pro-Flag series as shown before, *URA3*) in SD-URA medium and transformed with 100ng of pRS415 plasmids with/without HindIII digestion. 1/10 of transformants were spread onto uracil/leucine synthetic double drop-out plates (SC-Ura/SC-Leu). The colony numbers of both circular and linearized pRS415 transformants were counted after 2 days. The plasmid repair rate was calculated by normalizing the colony number of linearized pRS415 transformants with that of the circular pRS415 transformants.

**Table 4.1 Plasmid List**

| <b>Plasmid</b> | <b>Backbone</b> | <b>Description</b>                                  | <b>Source</b> |
|----------------|-----------------|---|---------------|
| <b>pWY084</b>  | pWY080          | <i>pRS416 ADH1pro-Flag</i>                          | this study    |
| <b>pWY088</b>  | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 FL</i>                  | this study    |
| <b>pWY157</b>  | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 Y149A</i>               | this study    |
| <b>pWY158</b>  | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 Y149N</i>               | this study    |
| <b>pWY159</b>  | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 F234Y</i>               | this study    |
| <b>pWY161</b>  | pWY105          | <i>pRS416-ADH1pro-FLAG-Set2 ΔB (Δ303-338)/Y149A</i> | this study    |
| <b>pWY162</b>  | pWY105          | <i>pRS416-ADH1pro-FLAG-Set2 ΔB (Δ303-338)/Y149N</i> | this study    |
| <b>pWY163</b>  | pWY105          | <i>pRS416-ADH1pro-FLAG-Set2 ΔB (Δ303-338)/F234Y</i> | this study    |
| <b>pWY164</b>  | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 Y149S</i>               | this study    |
| <b>pWY165</b>  | pWY084          | <i>pRS416-ADH1pro-FLAG-Set2 Y149F</i>               | this study    |
| <b>pWY105</b>  | pBL532          | <i>pBP-HFT-Set2 ΔB (Δ303-338)</i>                   | This study    |
| <b>pWY005</b>  | pBL532          | <i>pBP-HFT-Set2</i>                                 | this study    |
| <b>pWY166</b>  | pBL532          | <i>pBP-HTF-Y149A</i>                                | this study    |
| <b>pWY167</b>  | pBL532          | <i>pBP-HTF-Y149N</i>                                | this study    |
| <b>pWY168</b>  | pBL532          | <i>pBP-HTF-Y149F</i>                                | this study    |
| <b>pWY170</b>  | pBL532          | <i>pBP-HTF-F234Y</i>                                | this study    |
| <b>pBL532</b>  |                 | <i>pBacPAK-8-N-HisFlag-TEVsite</i>                  | Li Lab        |

Table 4.2 Stains List

| Name   | Parental strain | Genotype  | Source     |
|--------|-----------------|---|------------|
| YYW162 | w303            | <i>MATa ade2 can1 his3 leu2 lys2 met15 trp1 ura3 spt16-11</i>   | Li Lab     |
| YYW163 | w303            | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN</i>  | Li Lab     |
| YYW164 | w303            | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔRCO1::URA3</i>   | Li Lab     |
| YYW165 | YYW163          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN<br/>ΔRCO1::cre-HIS</i>   | Li Lab     |
| YYW809 | YYW162          | <i>MATa ade2 can1 his3 leu2 lys2 met15 trp1 ura3<br/>spt16-11+pWY084(pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>                 | this study |
| YYW810 | YYW163          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11<br/>ΔSET2::KAN+pWY084(pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>                      | this study |
| YYW811 | YYW163          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN+pWY088<br/>(pRS416-ADH1pro-FLAG-Set2 FL)</i>                           | this study |
| YYWS4  | YYW163          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11<br/>ΔSET2::KAN+pWY165(pRS416-ADH1pro-FLAG-Set2 Y149F)</i>                         | this study |
| YYWS5  | YYW163          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11<br/>ΔSET2::KAN+pWY159(pRS416-ADH1pro-FLAG-Set2 F234Y)</i>                         | this study |
| YYWS6  | YYW165          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN<br/>ΔRCO1::cre-HIS(c.kl)+pWY084(pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i> | this study |
| YYWS7  | YYW165          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN<br/>ΔRCO1::cre-HIS(c.kl)+pWY088(pRS416-ADH1pro-FLAG-Set2 FL)</i>       | this study |
| YYWS8  | YYW165          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN<br/>ΔRCO1::cre-HIS(c.kl)+pWY165(pRS416-ADH1pro-FLAG-Set2 Y149F)</i>    | this study |
| YYWS9  | YYW165          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11 ΔSET2::KAN<br/>ΔRCO1::cre-HIS(c.kl)+pWY159(pRS416-ADH1pro-FLAG-Set2 F234Y)</i>    | this study |
| YYWS83 | YYW164          | <i>MATa ade2 can1 his3 leu2 ura3 spt16-11<br/>ΔRCO1::URA3+pWY084(pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>                     | this study |
| YCR376 | BY4741          | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3</i>  | Li Lab     |
| YCR377 | YCR376          | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3<br/>ΔSET2::HPH</i>   | Li Lab     |
| YYWS10 | YCR377          | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3<br/>ΔSET2::HPH+pWY084 (pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>          | this study |
| YYWS11 | YCR377          | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3<br/>ΔSET2::HPH+pWY088 (pRS416-ADH1pro-FLAG-Set2 FL)</i>                | this study |

|        |        |  |              |
|--------|--------|--|--------------|
| YYWS12 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH+pWY165(pRS416-ADH1pro-FLAG-Set2 Y149F)</i>   | this study   |
| YYWS13 | YCR377 | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3 ΔSET2::HPH+pWY159(pRS416-ADH1pro-FLAG-Set2 F234Y)</i>   | this study   |
| YYWS15 | YCR376 | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3+pWY165(pRS416-ADH1pro-FLAG-Set2 Y149F)</i>              | this study   |
| YYWS16 | YCR376 | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 STE11 1840-HIS3+pWY159(pRS416-ADH1pro-FLAG-Set2 F234Y)</i>              | this study   |
| YYW005 | BY4741 | <i>MATa his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN</i>  | Li Lab       |
| YYW010 | BY4742 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN</i>  | Li Lab       |
| YYWS17 | BY4742 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0+pWY084 (pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>                      | this study   |
| YYWS18 | BY4742 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0+pWY165(pRS416-ADH1pro-FLAG-Set2 Y149F)</i>                          | this study   |
| YYWS19 | BY4742 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0+pWY159(pRS416-ADH1pro-FLAG-Set2 F234Y)</i>                          | this study   |
| YYWS20 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY088 (pRS416-ADH1pro-FLAG-Set2 FL)</i>                 | this study   |
| YYWS21 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY084 (pRS416-ADH1pro-APDT2-FLAG-CYC1ter)</i>           | this study   |
| YYWS22 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY165(pRS416-ADH1pro-FLAG-Set2 Y149F)</i>               | this study   |
| YYWS23 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY159(pRS416-ADH1pro-FLAG-Set2 F234Y)</i>               | this study   |
| YYWS23 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY157(pRS416-ADH1pro-FLAG-Set2 Y149A)</i>               | this study   |
| YYWS76 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY158(pRS416-ADH1pro-FLAG-Set2 Y149N)</i>               | this study   |
| YYWS77 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY164(pRS416-ADH1pro-FLAG-Set2 Y149S)</i>               | this study   |
| YYWS78 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY161(pRS416-ADH1pro-FLAG-Set2 ΔB (Δ303-338)/Y149A)</i> | this study   |
| YYWS79 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY162(pRS416-ADH1pro-FLAG-Set2 ΔB (Δ303-338)/Y149N)</i> | this study   |
| YYWS80 | YYW010 | <i>MATalpha his3Δ1 leu2Δ0 lys2Δ 0 ura3Δ0 ΔSET2::KAN+pWY163(pRS416-ADH1pro-FLAG-Set2 ΔB (Δ303-338)/F234Y)</i> | this study   |
| YYW157 |        | <i>MATaurs3-1 Leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [phi+] ess1-H164R</i>                             | Steven Hanes |

|                        |        |   |            |
|------------------------|--------|---|------------|
| <b>YYW160/<br/>161</b> |        | <i>MATa</i> <i>urs3-1 Leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [phi+]</i><br><i>ess1-H164R SET2A::HPH</i>   | this study |
| <b>YYWS27</b>          | YYW157 | <i>MATa</i> <i>urs3-1 Leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [phi+]</i><br><i>ess1-H164R+pWY084 (pRS416-ADH1pro-APDT2-FLAG-CYC1ter</i>                      | this study |
| <b>YYWS28</b>          | YYW160 | <i>MATa</i> <i>urs3-1 Leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [phi+]</i><br><i>ess1-H164R SET2A::HPH+pWY084</i><br><i>(pRS416-ADH1pro-APDT2-FLAG-CYC1ter</i> | this study |
| <b>YYWS29</b>          | YYW160 | <i>MATa</i> <i>urs3-1 Leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [phi+]</i><br><i>ess1-H164R SET2A::HPH+pWY088 (pRS416-ADH1pro-FLAG-Set2</i><br><i>FL)</i>      | this study |
| <b>YYWS81</b>          | YYW160 | <i>MATa</i> <i>urs3-1 Leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [phi+]</i><br><i>ess1-H164R SET2A::HPH+pWY165(pRS416-ADH1pro-FLAG-Set2</i><br><i>Y149F)</i>    | this study |
| <b>YYWS82</b>          | YYW160 | <i>MATa</i> <i>urs3-1 Leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [phi+]</i><br><i>ess1-H164R SET2A::HPH +pWY159(pRS416-ADH1pro-FLAG-Set2</i><br><i>F234Y)</i>   | this study |

## Chapter 5 SUMMARY AND FUTURE DIRECTIONS

### 5.1 Novel insights into H3K36 regulation

H3K36 methylation is an important histone “mark” ubiquitously presenting in eukaryotes from yeast to mammalian cells. Emerging lines of evidence suggest this enzyme is critical for development (Hu et al., 2010) and tumorigenesis (Duns et al., 2010; Newbold and Mokbel, 2010). In yeast, Set2 is responsible for conferring all three methylation states (mono-, di- and tri-) on histone H3 lysine 36 (Lee and Shilatifard, 2007; Strahl et al., 2002). Set2 mediated H3K36 methylation is tightly regulated.

In Chapter 2, we discovered a novel function of p-CTD interaction domain SRI. Besides recruiting Set2 to p-CTD, SRI also bound nucleosomal DNA which was essential for H3K36me<sub>3</sub> (Chapter 2, Fig 2.3 D and E). It is possible that SRI/DNA contact may increase the retention time for Set2 on its nucleosome substrates to facilitate tri-methyltransferase activity. Interestingly, structural prediction of SRI showed that its p-CTD and DNA interaction surfaces were mutually exclusive (Fig 3.7 C). Therefore, examining whether the deletion of DNA but not p-CTD interaction surface affects the H3K36me<sub>3</sub> status will give us insights into the detailed mechanism of how tri-methylation of H3K36 is regulated. We identified auto-inhibition domain (AID) in Set2 and demonstrate that the Set2-FL activity is fine-tuned by the opposing effects from AID and SRI domain (Fig 2.5E). Abnormal Set2 function by disrupting AID domain will cause synthetic growth phenotype with histone chaperon Spt16 (Fig 2.10 B).

In Chapter 4, we identified F/Y switch in yeast Set2 and successfully created both di-methylation and tri-methylation only mutants (Fig 4.3), which allow us to evaluate the state specific functions of H3K36me. We believe the catalytic mechanism that determines product specificity of Set2 is conserved, despite that the individual H3K36 methyl-state may have distinctive cellular functions. Therefore, we hope that the information we gain from the model system of yeast may help shed lights on how H3K36me is regulated in human, especially during development and tumorigenesis.

## 5.2 Explore the orchestration of PCAPs on p-CTD

The factors that have been found to associate with phosphorylated CTD tail include: elongation factors (Corden and Patturajan, 1997), mRNA capping enzymes (Cho et al., 1997; McCracken et al., 1997a), splicing enzymes (de la Mata and Kornblihtt, 2006), editing factors (Ryman et al., 2007), polyadenylation and 3' cleavage factors (McCracken et al., 1997b) transcription termination factors (Gudipati et al., 2008; Vasiljeva et al., 2008) (see review (Eick and Geyer, 2013)) and histone modifiers such as Set2 as extensively studied in this thesis. To our surprise, we found that excessive phosphorylation on CTD tail inhibited Set2 activity on nucleosome (Figure 3.8). We also showed that though Set2 had the ability to bind to longer p-CTD tail *in vitro*, its optimal substrate was three phosphorylated heptad repeats (Figure 3.9). Increasing repeats will decrease, instead of increase, its activity. This finding suggests that *in vivo* Set2 may bind best to p-CTD with a fix number of exposed phosphorylated sites and indicating a possible more organized recruitment of various PCAPs on p-CTD during transcription. We propose a model that CTD orchestrates the PCAPs by forming functional modules in which a subset of PCAPs are recruited by both available phosphorylation sites and cooperative binding with neighboring PCAPs. For instance, transcription termination factors Rtt103 and Pcf11 prefer longer (four instead of two) heptad peptide because two copies of each factors cooperatively bind to neighboring CTD repeats respectively (Lunde et al., 2010). Also in agreement with above speculation, yeast express CTD repeats longer than wild type also have growth defect under more under stressful conditions (15°C and 37°C), though cells can grow normally in 30°C (Liu et al, 2010). It is suggested that CTD with wild type length have denser packing and may better accommodate overall CTD-PCAPs interactions, and pre-established PCAP binding on the CTD subsequently affects the efficiency of interactions with additional stress response PCAPs. This result suggests that the globe CTD-PCAP organization rather than randomly loading of PCAPs is essential for CTD function.

In future study, we can take advantage of our engineered CTD system (Figure 3.4) to characterize how Set2 recognize its basic recognition motif (3 heptad repeats, as we showed in

the Chapter 3) along the long repeating domain of CTD. It is also interesting to identify the potential neighboring PCAPs of Set2 and how they work in concert to achieve precise temporal control of histone modifications during dynamic elongation process.

### **5.3 Looking for H3K36 methyl-state-specific readers.**

Histone PTMs can directly influence chromatin structure. For instance, acetylation on lysine residues can reduce the positive charge of histones, thereby weakening their interaction with negatively charged DNA and increasing nucleosome fluidity (Workman and Kingston, 1998). Moreover, the “histone code” hypothesis predicts that the modification marks on histone tails should provide binding sites for recruiting effector proteins which ultimately determine the functional outcome of certain PTMs (Jenuwein and Allis, 2001). The chromo domain, the PWWP domain and a hybrid Bromo-Zn-PWWP domain have been found to mediate the recognition of H3K36 methylation mark (Wen et al., 2014; Yun et al., 2011). The Chromo domain of Eaf3, the subunit of the Rpd3S HDAC complex, binds to histone methylated H3K36 with the help of the PHD domain of Rco1 subunit (Li et al., 2007b). In *Drosophila*, the male-specific lethal (MSL) complex which is required for X chromosome dosage compensation preferentially binds H3K36me3 marked nucleosomes through chromo domain containing MSL3 (Hosey and Brand, 2009; Larschan et al., 2007). In human, Eaf3 homolog MRG15 is recruited by H3K36me3 to modulates alternative splicing (Luco et al., 2010). Meanwhile PWWP domain containing proteins: DNMT3A (DNA methyltransferase), NSD family member proteins (H3K36 mono-, di- methyltransferase), MSH-6 (DNA mismatch recognition factor) are also shown to recognize methylated H3K36 to perform various function in cells (Dhayalan et al., 2010; Li et al., 2013; Wagner and Carpenter, 2012). DNMT3A and MSH-6 have been shown to specifically recognize H3K36me3 rather than other H3K36 methylation marks, rendering possibility that these factors can potentially mediate state-specific function of methylated H3K36 (Dhayalan et al., 2010; Li et al., 2013). In Chapter 4, we found that di- and tri- methylated H3K36 play distinctive roles in DNA damage repair pathway (Figure 4.8). Other works in human and yeast also reveal that H3K36me2 correlates with DSB processing and involves in the H3K36

methylation-dependent DNA damage repair (Fnu et al., 2011; Jha and Strahl, 2014). In the future study we plan to explore the detailed roles of specific H3K36 methylation in DNA damage repair pathway. With the help of high-throughput candidate based approach, the reader screening of state-specific histone methylation markers will be largely sped up (Kim et al., 2006). We hope the discovery of methyl-state histone modification readers will help us understand more detailed signals of histone code.

## Bibliography

1. Abu-Farha, M., Lambert, J.P., Al-Madhoun, A.S., Elisma, F., Skerjanc, I.S., and Figeys, D. (2008). The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase. *Mol Cell Proteomics* 7, 560-572.
2. Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular cell* 9, 1191-1200.
3. Al Sarakbi, W., Sasi, W., Jiang, W.G., Roberts, T., Newbold, R.F., and Mokbel, K. (2009). The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. *BMC Cancer* 9, 290.
4. Albert, A., Lavoie, S., and Vincent, M. (1999). A hyperphosphorylated form of RNA polymerase II is the major interphase antigen of the phosphoprotein antibody MPM-2 and interacts with the peptidyl-prolyl isomerase Pin1. *J Cell Sci* 112 ( Pt 15), 2493-2500.
5. An, S., Yeo, K.J., Jeon, Y.H., and Song, J.J. (2011). Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. *The Journal of biological chemistry* 286, 8369-8374.
6. Archambault, J., Chambers, R.S., Kobor, M.S., Ho, Y., Cartier, M., Bolotin, D., Andrews, B., Kane, C.M., and Greenblatt, J. (1997). An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 94, 14300-14305.
7. Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* 280, 7131-7134.
8. Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* 21, 381-395.
9. Bartkowiak, B., and Greenleaf, A.L. (2011). Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? *Transcription* 2, 115-119.
10. Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T., and Greenleaf, A.L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & development* 24, 2303-2316.
11. Baskaran, R., Dahmus, M.E., and Wang, J.Y. (1993). Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America* 90, 11167-11171.
12. Bataille, A.R., Jeronimo, C., Jacques, P.E., Laramée, L., Fortin, M.E., Forest, A., Bergeron, M., Hanes, S.D., and Robert, F. (2012). A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Molecular cell* 45, 158-170.
13. Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., Davies, H., *et al.* (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet* 45, 1479-1482.
14. Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A., and Schubeler, D. (2008). Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Molecular and cellular biology* 28, 3401-3409.
15. Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Bell, S.P., Imhof, A., *et al.* (2007). Localized H3K36 methylation states define histone H4K16 acetylation during

transcriptional elongation in *Drosophila*. *Embo J* 26, 4974-4984.

16. Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science* 301, 1090-1093.
17. Biswas, D., Dutta-Biswas, R., Mitra, D., Shibata, Y., Strahl, B.D., Formosa, T., and Stillman, D.J. (2006). Opposing roles for Set2 and yFACT in regulating TBP binding at promoters. *The EMBO journal* 25, 4479-4489.
18. Biswas, D., Takahata, S., and Stillman, D.J. (2008a). Different genetic functions for the Rpd3(L) and Rpd3(S) complexes suggest competition between NuA4 and Rpd3(S). *Molecular and cellular biology* 28, 4445-4458.
19. Biswas, D., Takahata, S., Xin, H., Dutta-Biswas, R., Yu, Y., Formosa, T., and Stillman, D.J. (2008b). A role for Chd1 and Set2 in negatively regulating DNA replication in *Saccharomyces cerevisiae*. *Genetics* 178, 649-659.
20. Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* 44, 148-156.
21. Botuyan, M.V., Lee, J., Ward, I.M., Kim, J.E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* 127, 1361-1373.
22. Brown, M.A., Sims, R.J., 3rd, Gottlieb, P.D., and Tucker, P.W. (2006). Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol Cancer* 5, 26.
23. Brugarolas, J. (2014). Molecular genetics of clear-cell renal cell carcinoma. *J Clin Oncol* 32, 1968-1976.
24. Buratowski, S. (2003). The CTD code. *Nat Struct Biol* 10, 679-680.
25. Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Molecular cell* 36, 541-546.
26. Burgess, R.J., and Zhang, Z. (2013). Histone chaperones in nucleosome assembly and human disease. *Nat Struct Mol Biol* 20, 14-22.
27. Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-angstrom resolution: Implications for the initiation of transcription. *Proceedings of the National Academy of Sciences of the United States of America* 100, 6969-6973.
28. Canzio, D., Liao, M., Naber, N., Pate, E., Larson, A., Wu, S., Marina, D.B., Garcia, J.F., Madhani, H.D., Cooke, R., *et al.* (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature* 496, 377-381.
29. Carey, M., Li, B., and Workman, J.L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Molecular cell* 24, 481-487.
30. Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005a). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592.
31. Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005b). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592.
32. Carty, S.M., Goldstrohm, A.C., Sune, C., Garcia-Blanco, M.A., and Greenleaf, A.L. (2000). Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* 97, 9015-9020.
33. Chambers, R.S., and Dahmus, M.E. (1994). Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. *The Journal of biological chemistry* 269,

26243-26248.

34. Chandrasekharan, M.B., Huang, F., and Sun, Z.W. (2009). Ubiquitination of histone H2B regulates chromatin dynamics by enhancing nucleosome stability. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 16686-16691.
35. Chang, A., Cheang, S., Espanel, X., and Sudol, M. (2000). Rsp5 WW domains interact directly with the carboxyl-terminal domain of RNA polymerase II. *The Journal of biological chemistry* *275*, 20562-20571.
36. Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* *318*, 1780-1782.
37. Cheung, V., Chua, G., Batada, N.N., Landry, C.R., Michnick, S.W., Hughes, T.R., and Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS Biol* *6*, e277.
38. Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes & development* *15*, 3319-3329.
39. Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & development* *11*, 3319-3326.
40. Chu, Y., Simic, R., Warner, M.H., Arndt, K.M., and Prelich, G. (2007). Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *Embo J* *26*, 4646-4656.
41. Chu, Y., Sutton, A., Sternglanz, R., and Prelich, G. (2006). The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. *Molecular and cellular biology* *26*, 3029-3038.
42. Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem* *78*, 273-304.
43. Collins, R.E., Tachibana, M., Tamaru, H., Smith, K.M., Jia, D., Zhang, X., Selker, E.U., Shinkai, Y., and Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *The Journal of biological chemistry* *280*, 5563-5570.
44. Comer, F.I., and Hart, G.W. (2001). Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* *40*, 7845-7852.
45. Corden, J., David, T., and Fisher, J. (1995). Determination of the curvatures and bending strains in open trileaflet heart valves. *Proc Inst Mech Eng H* *209*, 121-128.
46. Corden, J.L. (2013a). RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chem Rev* *113*, 8423-8455.
47. Corden, J.L. (2013b). RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chem Rev* *113*, 8423-8455.
48. Corden, J.L., and Patturajan, M. (1997). A CTD function linking transcription to splicing. *Trends Biochem Sci* *22*, 413-416.
49. Couture, J.F., Dirk, L.M., Brunzelle, J.S., Houtz, R.L., and Trievel, R.C. (2008). Structural origins for the product specificity of SET domain protein methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 20659-20664.
50. Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.
51. Creighton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A.,

- Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 21931-21936.
52. de la Mata, M., and Kornblihtt, A.R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nature structural & molecular biology* *13*, 973-980.
53. Del Rizzo, P.A., Couture, J.F., Dirk, L.M., Strunk, B.S., Roiko, M.S., Brunzelle, J.S., Houtz, R.L., and Trievel, R.C. (2010). SET7/9 catalytic mutants reveal the role of active site water molecules in lysine multiple methylation. *The Journal of biological chemistry* *285*, 31849-31858.
54. Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* *399*, 491-496.
55. Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* *285*, 26114-26120.
56. Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* *6*, 227.
57. Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. *Science* *315*, 1405-1408.
58. Du, H.N., and Briggs, S.D. (2010). A nucleosome surface formed by histone H4, H2A, and H3 residues is needed for proper histone H3 Lys36 methylation, histone acetylation, and repression of cryptic transcription. *J Biol Chem* *285*, 11704-11713.
59. Du, H.N., Fingerman, I.M., and Briggs, S.D. (2008). Histone H3 K36 methylation is mediated by a trans-histone methylation pathway involving an interaction between Set2 and histone H4. *Genes Dev* *22*, 2786-2798.
60. Duns, G., van den Berg, E., van Duivenbode, I., Osinga, J., Hollema, H., Hofstra, R.M., and Kok, K. (2010). Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer research* *70*, 4287-4291.
61. Eberharter, A., John, S., Grant, P.A., Utley, R.T., and Workman, J.L. (1998). Identification and analysis of yeast nucleosomal histone acetyltransferase complexes. *Methods* *15*, 315-321.
62. Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *Embo J* *27*, 406-420.
63. Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* *318*, 1777-1779.
64. Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S. (2010). The Integrator Complex Recognizes a New Double Mark on the RNA Polymerase II Carboxyl-terminal Domain. *Journal of Biological Chemistry* *285*, 20564-20569.
65. Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* *113*, 8456-8490.
66. Engen, J.R. (2009). Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. *Anal Chem* *81*, 7870-7875.
67. Eom, G.H., Kim, K.B., Kim, J.H., Kim, J.Y., Kim, J.R., Kee, H.J., Kim, D.W., Choe, N., Park, H.J., Son, H.J., *et al.* (2011). Histone methyltransferase SETD3 regulates muscle differentiation. *The Journal of biological chemistry* *286*, 34733-34742.

68. Fabrega, C., Hausmann, S., Shen, V., Shuman, S., and Lima, C.D. (2004). Structure and mechanism of mRNA cap (guanine-N7) methyltransferase. *Molecular cell* *13*, 77-89.
69. Fang, J., Feng, Q., Ketel, C.S., Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., Simon, J.A., and Zhang, Y. (2002). Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr Biol* *12*, 1086-1099.
70. Fang, J., Hogan, G.J., Liang, G., Lieb, J.D., and Zhang, Y. (2007). The *Saccharomyces cerevisiae* histone demethylase Jhd1 fine-tunes the distribution of H3K36me2. *Molecular and cellular biology* *27*, 5055-5065.
71. Feaver, W.J., Gileadi, O., Li, Y., and Kornberg, R.D. (1991). CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* *67*, 1223-1230.
72. Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* *12*, 1052-1058.
73. Filipescu, D., Szenker, E., and Almouzni, G. (2013). Developmental roles of histone H3 variants and their chaperones. *Trends in genetics : TIG* *29*, 630-640.
74. Fleming, A.B., Kao, C.F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Molecular cell* *31*, 57-66.
75. Flynn, R.A., and Chang, H.Y. (2012). Active chromatin and noncoding RNAs: an intimate relationship. *Curr Opin Genet Dev* *22*, 172-178.
76. Fnu, S., Williamson, E.A., De Haro, L.P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S.H., Nickoloff, J.A., and Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc Natl Acad Sci U S A* *108*, 540-545.
77. Fontebasso, A.M., Schwartzenrubler, J., Khuong-Quang, D.A., Liu, X.Y., Sturm, D., Korshunov, A., Jones, D.T., Witt, H., Kool, M., Albrecht, S., *et al.* (2013). Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta Neuropathol* *125*, 659-669.
78. Frederiks, F., Tzouros, M., Oudgenoeg, G., van Welsem, T., Fornerod, M., Krijgsveld, J., and van Leeuwen, F. (2008). Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. *Nat Struct Mol Biol* *15*, 550-557.
79. Fuchs, S.M., Kizer, K.O., Braberg, H., Krogan, N.J., and Strahl, B.D. (2012). RNA polymerase II carboxyl-terminal domain phosphorylation regulates protein stability of the Set2 methyltransferase and histone H3 di- and trimethylation at lysine 36. *The Journal of biological chemistry* *287*, 3249-3256.
80. Gao, Y.G., Yang, H., Zhao, J., Jiang, Y.J., and Hu, H.Y. (2014). Autoinhibitory Structure of the WW Domain of HYPB/SETD2 Regulates Its Interaction with the Proline-Rich Region of Huntingtin. *Structure*.
81. Gemmill, T.R., Wu, X., and Hanes, S.D. (2005). Vanishingly low levels of Ess1 prolyl-isomerase activity are sufficient for growth in *Saccharomyces cerevisiae*. *The Journal of biological chemistry* *280*, 15510-15517.
82. Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIF-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Molecular and cellular biology* *29*, 5455-5464.
83. Goldman, J.A., Garlick, J.D., and Kingston, R.E. (2010). Chromatin remodeling by imitation switch (ISWI) class ATP-dependent remodelers is stimulated by histone variant H2A.Z. *The Journal of biological chemistry* *285*, 4645-4651.
84. Gonzalo, S., Garcia-Cao, M., Fraga, M.F., Schotta, G., Peters, A.H., Cotter, S.E., Eguia, R., Dean, D.C., Esteller, M., Jenuwein, T., *et al.* (2005). Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin.

Nature cell biology 7, 420-428.

85. Govind, C.K., Qiu, H., Ginsburg, D.S., Ruan, C., Hofmeyer, K., Hu, C., Swaminathan, V., Workman, J.L., Li, B., and Hinnebusch, A.G. (2010). Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Mol Cell* 39, 234-246.
86. Gregory, G.D., Vakoc, C.R., Rozovskaia, T., Zheng, X., Patel, S., Nakamura, T., Canaani, E., and Blobel, G.A. (2007). Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Molecular and cellular biology* 27, 8466-8479.
87. Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nature structural & molecular biology* 15, 786-794.
88. Hajdu, I., Ciccia, A., Lewis, S.M., and Elledge, S.J. (2011). Wolf-Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. *Proceedings of the National Academy of Sciences of the United States of America* 108, 13130-13134.
89. Hanes, S.D., Shank, P.R., and Bostian, K.A. (1989). Sequence and mutational analysis of ESS1, a gene essential for growth in *Saccharomyces cerevisiae*. *Yeast* 5, 55-72.
90. Hani, J., Stumpf, G., and Domdey, H. (1995). PTF1 encodes an essential protein in *Saccharomyces cerevisiae*, which shows strong homology with a new putative family of PPlases. *FEBS letters* 365, 198-202.
91. Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* 21, 396-420.
92. Hausmann, S., Koiwa, H., Krishnamurthy, S., Hampsey, M., and Shuman, S. (2005). Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. *The Journal of biological chemistry* 280, 37681-37688.
93. Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., *et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The EMBO journal* 31, 2784-2797.
94. Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (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. *The Journal of biological chemistry* 268, 305-314.
95. Hosey, A.M., and Brand, M. (2009). Chromodomain-mediated spreading on active genes. *Nature structural & molecular biology* 16, 11-13.
96. Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & development* 26, 2119-2137.
97. Hsin, J.P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* 334, 683-686.
98. Hu, M., Sun, X.J., Zhang, Y.L., Kuang, Y., Hu, C.Q., Wu, W.L., Shen, S.H., Du, T.T., Li, H., He, F., *et al.* (2010). Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc Natl Acad Sci U S A* 107, 2956-2961.
99. Huang, Y., Fang, J., Bedford, M.T., Zhang, Y., and Xu, R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* 312, 748-751.
100. Hudlebusch, H.R., Santoni-Rugiu, E., Simon, R., Ralfkiaer, E., Rossing, H.H., Johansen, J.V., Jorgensen, M., Sauter, G., and Helin, K. (2011a). The histone methyltransferase and putative oncoprotein MMSET is overexpressed in a large variety of human tumors. *Clinical cancer research : an official journal of the American Association for Cancer*

Research 17, 2919-2933.

101. Hudlebusch, H.R., Skotte, J., Santoni-Rugiu, E., Zimling, Z.G., Lees, M.J., Simon, R., Sauter, G., Rota, R., De Ioris, M.A., Quarto, M., *et al.* (2011b). MMSET is highly expressed and associated with aggressiveness in neuroblastoma. *Cancer research* 71, 4226-4235.
102. Jaehning, J.A. (2010). The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochimica et biophysica acta* 1799, 379-388.
103. Jelinic, P., Pellegrino, J., and David, G. (2011). A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Mol Cell Biol* 31, 54-62.
104. Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074-1080.
105. Jha, D.K., and Strahl, B.D. (2014). An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair. *Nature communications* 5, 3965.
106. Jorgensen, S., Schotta, G., and Sorensen, C.S. (2013). Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic acids research* 41, 2797-2806.
107. Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Molecular cell* 20, 971-978.
108. Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301, 1096-1099.
109. Kassambara, A., Klein, B., and Moreaux, J. (2009). MMSET is overexpressed in cancers: link with tumor aggressiveness. *Biochem Biophys Res Commun* 379, 840-845.
110. Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *The Journal of biological chemistry* 268, 10416-10424.
111. Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., *et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123, 593-605.
112. Keogh, M.C., Podolny, V., and Buratowski, S. (2003). Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Molecular and cellular biology* 23, 7005-7018.
113. Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 17, 1279-1286.
114. Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* 7, 397-403.
115. Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeja, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517-522.
116. Kim, T., and Buratowski, S. (2007). Two *Saccharomyces cerevisiae* JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. *J Biol Chem* 282, 20827-20835.
117. Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology* 25, 3305-3316.
118. Klose, R.J., Gardner, K.E., Liang, G., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2007). Demethylation of histone H3K36 and H3K9 by Rph1: a vestige of an H3K9 methylation system in *Saccharomyces cerevisiae*? *Mol Cell Biol* 27, 3951-3961.
119. Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. *Nat*

Rev Genet 7, 715-727.

120. Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* 41, 376-381.
121. Kornberg, R.D. (1977). Structure of chromatin. *Annu Rev Biochem* 46, 931-954.
122. Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.
123. Krishnamurthy, S., Ghazy, M.A., Moore, C., and Hampsey, M. (2009). Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. *Molecular and cellular biology* 29, 2925-2934.
124. Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., *et al.* (2003). Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Molecular and cellular biology* 23, 4207-4218.
125. Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes & development* 26, 1891-1896.
126. Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Luring, J., Xi, Y., Park, B.H., Shi, X., Garcia, B.A., *et al.* (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Molecular cell* 44, 609-620.
127. Lachner, M., O'Sullivan, R.J., and Jenuwein, T. (2003). An epigenetic road map for histone lysine methylation. *J Cell Sci* 116, 2117-2124.
128. Lange, M., Kaynak, B., Forster, U.B., Tonjes, M., Fischer, J.J., Grimm, C., Schlesinger, J., Just, S., Dunkel, I., Krueger, T., *et al.* (2008). Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev* 22, 2370-2384.
129. Larschan, E., Alekseyenko, A.A., Gortchakov, A.A., Peng, S., Li, B., Yang, P., Workman, J.L., Park, P.J., and Kuroda, M.I. (2007). MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* 28, 121-133.
130. Larschan, E., Bishop, E.P., Kharchenko, P.V., Core, L.J., Lis, J.T., Park, P.J., and Kuroda, M.I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature* 471, 115-118.
131. Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021-1036.
132. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495-501.
133. Lee, J.M., and Greenleaf, A.L. (1991). CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of *Saccharomyces cerevisiae*. *Gene Expr* 1, 149-167.
134. Lee, J.S., and Shilatifard, A. (2007). A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat Res* 618, 130-134.
135. Lee, S.H., Oshige, M., Durant, S.T., Rasila, K.K., Williamson, E.A., Ramsey, H., Kwan, L., Nickoloff, J.A., and Hromas, R. (2005). The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. *Proceedings of the National Academy of Sciences of the United States of America* 102, 18075-18080.
136. Li, B., Carey, M., and Workman, J.L. (2007a). The role of chromatin during transcription. *Cell* 128, 707-719.
137. Li, B., Gogol, M., Carey, M., Lee, D., Seidel, C., and Workman, J.L. (2007b). Combined action of PHD and

- chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science* 316, 1050-1054.
138. Li, B., Gogol, M., Carey, M., Pattenden, S.G., Seidel, C., and Workman, J.L. (2007c). Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev* 21, 1422-1430.
139. Li, B., Howe, L., Anderson, S., Yates, J.R., 3rd, and Workman, J.L. (2003a). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *The Journal of biological chemistry* 278, 8897-8903.
140. Li, B., Howe, L., Anderson, S., Yates, J.R., 3rd, and Workman, J.L. (2003b). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* 278, 8897-8903.
141. Li, B., Jackson, J., Simon, M.D., Fleharty, B., Gogol, M., Seidel, C., Workman, J.L., and Shilatifard, A. (2009a). Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *J Biol Chem* 284, 7970-7976.
142. Li, B., and Reese, J.C. (2001). Ssn6-Tup1 regulates RNR3 by positioning nucleosomes and affecting the chromatin structure at the upstream repression sequence. *J Biol Chem* 276, 33788-33797.
143. Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. *Cell* 153, 590-600.
144. Li, M., Phatnani, H.P., Guan, Z., Sage, H., Greenleaf, A.L., and Zhou, P. (2005). Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proceedings of the National Academy of Sciences of the United States of America* 102, 17636-17641.
145. Li, Y., Trojer, P., Xu, C.F., Cheung, P., Kuo, A., Drury, W.J., 3rd, Qiao, Q., Neubert, T.A., Xu, R.M., Gozani, O., *et al.* (2009b). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem* 284, 34283-34295.
146. Liao, S.M., Zhang, J., Jeffery, D.A., Koleske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., van Vuuren, H.J., and Young, R.A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* 374, 193-196.
147. Lin, L.J., Minard, L.V., Johnston, G.C., Singer, R.A., and Schultz, M.C. (2010). Asf1 can promote trimethylation of H3 K36 by Set2. *Mol Cell Biol* 30, 1116-1129.
148. Liu, P., Kenney, J.M., Stiller, J.W., and Greenleaf, A.L. (2010). Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol* 27, 2628-2641.
149. Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Molecular and cellular biology* 24, 1721-1735.
150. Lu, H., Flores, O., Weinmann, R., and Reinberg, D. (1991). The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc Natl Acad Sci U S A* 88, 10004-10008.
151. Lu, P.J., Zhou, X.Z., Shen, M., and Lu, K.P. (1999). Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* 283, 1325-1328.
152. Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* 327, 996-1000.
153. Luger, K., and Richmond, T.J. (1998). DNA binding within the nucleosome core. *Curr Opin Struct Biol* 8, 33-40.
154. Luk, E., Ranjan, A., Fitzgerald, P.C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. *Cell* 143, 725-736.
155. Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A.,

- and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nature structural & molecular biology* 17, 1195-+.
156. Ma, Z., Atencio, D., Barnes, C., DeFiglio, H., and Hanes, S.D. (2012). Multiple roles for the Ess1 prolyl isomerase in the RNA polymerase II transcription cycle. *Mol Cell Biol* 32, 3594-3607.
157. Macdonald, N., Welburn, J.P., Noble, M.E., Nguyen, A., Yaffe, M.B., Clynes, D., Moggs, J.G., Orphanides, G., Thomson, S., Edmunds, J.W., *et al.* (2005). Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. *Mol Cell* 20, 199-211.
158. Macias, M.J., Wiesner, S., and Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS letters* 513, 30-37.
159. Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nat Struct Biol* 10, 882-891.
160. Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499-506.
161. Mar, B.G., Bullinger, L.B., McLean, K.M., Grauman, P.V., Harris, M.H., Stevenson, K., Neuberg, D.S., Sinha, A.U., Sallan, S.E., Silverman, L.B., *et al.* (2014). Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. *Nature communications* 5, 3469.
162. Marango, J., Shimoyama, M., Nishio, H., Meyer, J.A., Min, D.J., Sirulnik, A., Martinez-Martinez, Y., Chesi, M., Bergsagel, P.L., Zhou, M.M., *et al.* (2008). The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood* 111, 3145-3154.
163. Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *The Journal of biological chemistry* 271, 27176-27183.
164. Mayer, A., Heidemann, M., Lidschreiber, M., Schrieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336, 1723-1725.
165. McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & development* 11, 3306-3318.
166. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357-361.
167. Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. *Genes & development* 19, 1401-1415.
168. Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* 37, 1090-1097.
169. Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 303, 343-348.
170. Morris, D.P., and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *The Journal of biological chemistry* 275, 39935-39943.
171. Morris, D.P., Phatnani, H.P., and Greenleaf, A.L. (1999). Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-End formation. *The Journal of biological chemistry* 274, 31583-31587.
172. Morris, S.A., Rao, B., Garcia, B.A., Hake, S.B., Diaz, R.L., Shabanowitz, J., Hunt, D.F., Allis, C.D., Lieb, J.D., and Strahl, B.D. (2007). Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. *J*

Biol Chem 282, 7632-7640.

173. Morris, S.A., Shibata, Y., Noma, K., Tsukamoto, Y., Warren, E., Temple, B., Grewal, S.I., and Strahl, B.D. (2005). Histone H3 K36 methylation is associated with transcription elongation in *Schizosaccharomyces pombe*. *Eukaryot Cell* 4, 1446-1454.

174. Mumberg, D., Muller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156, 119-122.

175. Myers, J.K., Morris, D.P., Greenleaf, A.L., and Oas, T.G. (2001). Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1. *Biochemistry* 40, 8479-8486.

176. Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et biophysica acta* 1809, 34-45.

177. Nelson, C.J., Santos-Rosa, H., and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* 126, 905-916.

178. Newbold, R.F., and Mokbel, K. (2010). Evidence for a tumour suppressor function of SETD2 in human breast cancer: a new hypothesis. *Anticancer research* 30, 3309-3311.

179. Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* 11, 709-719.

180. Nguyen, A.T., and Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. *Genes & development* 25, 1345-1358.

181. Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., *et al.* (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Molecular cell* 9, 1201-1213.

182. Ooga, M., Inoue, A., Kageyama, S., Akiyama, T., Nagata, M., and Aoki, F. (2008). Changes in H3K79 methylation during preimplantation development in mice. *Biol Reprod* 78, 413-424.

183. Pai, C.C., Deegan, R.S., Subramanian, L., Gal, C., Sarkar, S., Blaikley, E.J., Walker, C., Hulme, L., Bernhard, E., Codlin, S., *et al.* (2014). A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. *Nature communications* 5, 4091.

184. Pattenden, S.G., Gogol, M.M., and Workman, J.L. (2010). Features of cryptic promoters and their varied reliance on bromodomain-containing factors. *PLoS One* 5, e12927.

185. Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* 125, 703-717.

186. Pei, H., Zhang, L., Luo, K., Qin, Y., Chesi, M., Fei, F., Bergsagel, P.L., Wang, L., You, Z., and Lou, Z. (2011). MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature* 470, 124-128.

187. Pfister, S.X., Ahrabi, S., Zalmas, L.P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday, T., Legube, G., La Thangue, N.B., Porter, A.C., *et al.* (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell reports* 7, 2006-2018.

188. Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & development* 20, 2922-2936.

189. Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A.,

- Herbolsheimer, E., *et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* *122*, 517-527.
190. Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS genetics* *8*, e1002717.
191. Prelich, G., and Winston, F. (1993). Mutations that suppress the deletion of an upstream activating sequence in yeast: involvement of a protein kinase and histone H3 in repressing transcription *in vivo*. *Genetics* *135*, 665-676.
192. Psathas, J.N., Zheng, S., Tan, S., and Reese, J.C. (2009). Set2-dependent K36 methylation is regulated by novel intratail interactions within H3. *Mol Cell Biol* *29*, 6413-6426.
193. Pufall, M.A., and Graves, B.J. (2002). Autoinhibitory domains: modular effectors of cellular regulation. *Annu Rev Cell Dev Biol* *18*, 421-462.
194. Qian, C., Wang, X., Manzur, K., Sachchidanand, Farooq, A., Zeng, L., Wang, R., and Zhou, M.M. (2006). Structural insights of the specificity and catalysis of a viral histone H3 lysine 27 methyltransferase. *Journal of molecular biology* *359*, 86-96.
195. Qian, C., and Zhou, M.M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci* *63*, 2755-2763.
196. Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Molecular cell* *33*, 752-762.
197. Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., and Lewis, B.A. (2012). Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription *in vitro* and *in vivo*. *The Journal of biological chemistry* *287*, 23549-23561.
198. Rao, B., Shibata, Y., Strahl, B.D., and Lieb, J.D. (2005). Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Molecular and cellular biology* *25*, 9447-9459.
199. Rayasam, G.V., Wendling, O., Angrand, P.O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *The EMBO journal* *22*, 3153-3163.
200. Regha, K., Sloane, M.A., Huang, R., Pauler, F.M., Warczok, K.E., Melikant, B., Radolf, M., Martens, J.H., Schotta, G., Jenuwein, T., *et al.* (2007). Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. *Molecular cell* *27*, 353-366.
201. Reid, J.L., Moqtaderi, Z., and Struhl, K. (2004). Eaf3 regulates the global pattern of histone acetylation in *Saccharomyces cerevisiae*. *Molecular and cellular biology* *24*, 757-764.
202. Rosati, R., La Starza, R., Veronese, A., Aventin, A., Schwienbacher, C., Vallespi, T., Negrini, M., Martelli, M.F., and Mecucci, C. (2002). NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). *Blood* *99*, 3857-3860.
203. Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell* *25*, 15-30.
204. Ryman, K., Fong, N., Bratt, E., Bentley, D.L., and Ohman, M. (2007). The C-terminal domain of RNA Pol II helps ensure that editing precedes splicing of the GluR-B transcript. *Rna* *13*, 1071-1078.
205. Sanders, S.L., Portoso, M., Mata, J., Bahler, J., Allshire, R.C., and Kouzarides, T. (2004). Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* *119*, 603-614.
206. Santisteban, M.S., Hang, M., and Smith, M.M. (2011). Histone variant H2A.Z and RNA polymerase II

- transcription elongation. *Molecular and cellular biology* 31, 1848-1860.
207. Schlichter, A., and Cairns, B.R. (2005). Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. *EMBO J* 24, 1222-1231.
208. Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & development* 18, 1251-1262.
209. Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A., *et al.* (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes & development* 22, 2048-2061.
210. Schulze, J.M., Jackson, J., Nakanishi, S., Gardner, J.M., Hentrich, T., Haug, J., Johnston, M., Jaspersen, S.L., Kobor, M.S., and Shilatifard, A. (2009). Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. *Molecular cell* 35, 626-641.
211. Schwer, B., Sanchez, A.M., and Shuman, S. (2012). Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America* 109, 18024-18029.
212. Schwer, B., and Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell* 43, 311-318.
213. Shi, Y., and Whetstine, J.R. (2007). Dynamic regulation of histone lysine methylation by demethylases. *Molecular cell* 25, 1-14.
214. Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* 81, 65-95.
215. Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844-847.
216. Silva, A.C., Xu, X., Kim, H.S., Fillingham, J., Kislinger, T., Mennella, T.A., and Keogh, M.C. (2012). The replication-independent histone H3-H4 chaperones HIR, ASF1, and RTT106 co-operate to maintain promoter fidelity. *The Journal of biological chemistry* 287, 1709-1718.
217. Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nature structural & molecular biology* 19, 884-892.
218. Sogaard, T.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *The Journal of biological chemistry* 282, 14113-14120.
219. Southall, S.M., Wong, P.S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Molecular cell* 33, 181-191.
220. Sterner, D.E., Lee, J.M., Hardin, S.E., and Greenleaf, A.L. (1995). The Yeast Carboxyl-Terminal Repeat Domain Kinase Ctdk-I Is a Divergent Cyclin-Cyclin-Dependent Kinase Complex. *Molecular and cellular biology* 15, 5716-5724.
221. Stiller, J.W., and Cook, M.S. (2004). Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell* 3, 735-740.
222. Stiller, J.W., McConaughy, B.L., and Hall, B.D. (2000). Evolutionary complementation for polymerase II CTD function. *Yeast* 16, 57-64.

223. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41-45.
224. Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., *et al.* (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Molecular and cellular biology* *22*, 1298-1306.
225. Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* *123*, 1213-1226.
226. Suka, N., Luo, K., and Grunstein, M. (2002). Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nat Genet* *32*, 378-383.
227. Sun, X.J., Wei, J., Wu, X.Y., Hu, M., Wang, L., Wang, H.H., Zhang, Q.H., Chen, S.J., Huang, Q.H., and Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *The Journal of biological chemistry* *280*, 35261-35271.
228. Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., *et al.* (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & development* *16*, 1779-1791.
229. Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. *Nature reviews Molecular cell biology* *11*, 264-275.
230. Tardat, M., Murr, R., Herceg, Z., Sardet, C., and Julien, E. (2007). PR-Set7-dependent lysine methylation ensures genome replication and stability through S phase. *The Journal of cell biology* *179*, 1413-1426.
231. Taylor, W.R., Xiao, B., Gamblin, S.J., and Lin, K. (2003). A knot or not a knot? SETting the record 'straight' on proteins. *Comput Biol Chem* *27*, 11-15.
232. Trievel, R.C., Beach, B.M., Dirk, L.M., Houtz, R.L., and Hurley, J.H. (2002). Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* *111*, 91-103.
233. van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* *109*, 745-756.
234. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature structural & molecular biology* *15*, 795-804.
235. Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., and Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* *489*, 452-455.
236. Venkatesh, S., and Workman, J.L. (2013). Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. *Wiley interdisciplinary reviews Developmental biology* *2*, 685-700.
237. Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Biol* *7*, 639-643.
238. Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* *131*, 58-69.
239. Vojnic, E., Simon, B., Strahl, B.D., Sattler, M., and Cramer, P. (2006). Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *The Journal of biological chemistry* *281*, 13-15.

240. Volkel, P., and Angrand, P.O. (2007). The control of histone lysine methylation in epigenetic regulation. *Biochimie* 89, 1-20.
241. Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* 13, 115-126.
242. Walter, W., Clynes, D., Tang, Y., Marmorstein, R., Mellor, J., and Berger, S.L. (2008). 14-3-3 interaction with histone H3 involves a dual modification pattern of phosphoacetylation. *Mol Cell Biol* 28, 2840-2849.
243. Wang, G.G., Cai, L., Pasillas, M.P., and Kamps, M.P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat Cell Biol* 9, 804-812.
244. Wang, Y., and Jia, S. (2009). Degrees make all the difference: the multifunctionality of histone H4 lysine 20 methylation. *Epigenetics* 4, 273-276.
245. Weber, C.M., and Henikoff, S. (2014). Histone variants: dynamic punctuation in transcription. *Genes & development* 28, 672-682.
246. Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., *et al.* (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* 508, 263-268.
247. West, M.L., and Corden, J.L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140, 1223-1233.
248. Winsor, T.S., Bartkowiak, B., Bennett, C.B., and Greenleaf, A.L. (2013). A DNA damage response system associated with the phosphoCTD of elongating RNA polymerase II. *PLoS One* 8, e60909.
249. Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol* 2, a000596.
250. Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* 67, 545-579.
251. Wu, S., Wang, W., Kong, X., Congdon, L.M., Yokomori, K., Kirschner, M.W., and Rice, J.C. (2010). Dynamic regulation of the PR-Set7 histone methyltransferase is required for normal cell cycle progression. *Genes & development* 24, 2531-2542.
252. Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442, 86-90.
253. Xiao, B., Jing, C., Kelly, G., Walker, P.A., Muskett, F.W., Frenkiel, T.A., Martin, S.R., Sarma, K., Reinberg, D., Gambelin, S.J., *et al.* (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes & development* 19, 1444-1454.
254. Xiao, B., Jing, C., Wilson, J.R., Walker, P.A., Vasisht, N., Kelly, G., Howell, S., Taylor, I.A., Blackburn, G.M., and Gambelin, S.J. (2003a). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* 421, 652-656.
255. Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003b). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes & development* 17, 654-663.
256. Xu, Y.X., Hirose, Y., Zhou, X.Z., Lu, K.P., and Manley, J.L. (2003). Pin1 modulates the structure and function of human RNA polymerase II. *Genes & development* 17, 2765-2776.
257. Yaffe, M.B., Schutkowski, M., Shen, M., Zhou, X.Z., Stukenberg, P.T., Rahfeld, J.U., Xu, J., Kuang, J., Kirschner, M.W., Fischer, G., *et al.* (1997). Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science* 278, 1957-1960.

258. Yang, H., Pesavento, J.J., Starnes, T.W., Cryderman, D.E., Wallrath, L.L., Kelleher, N.L., and Mizzen, C.A. (2008). Preferential dimethylation of histone H4 lysine 20 by Suv4-20. *The Journal of biological chemistry* **283**, 12085-12092.
259. Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* **22**, 3422-3434.
260. Youdell, M.L., Kizer, K.O., Kisseleva-Romanova, E., Fuchs, S.M., Duro, E., Strahl, B.D., and Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Molecular and cellular biology* **28**, 4915-4926.
261. Yuan, W., Xie, J., Long, C., Erdjument-Bromage, H., Ding, X., Zheng, Y., Tempst, P., Chen, S., Zhu, B., and Reinberg, D. (2009). Heterogeneous nuclear ribonucleoprotein L is a subunit of human KMT3a/Set2 complex required for H3 Lys-36 trimethylation activity in vivo. *The Journal of biological chemistry* **284**, 15701-15707.
262. Yun, M., Ruan, C., Huh, J.W., and Li, B. (2012). Reconstitution of modified chromatin templates for in vitro functional assays. *Methods Mol Biol* **833**, 237-253.
263. Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. *Cell Res* **21**, 564-578.
264. Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**, 258-262.
265. Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nature structural & molecular biology* **20**, 259-266.
266. Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **21**, 1273-1283.
267. Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodriguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012a). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *The Journal of biological chemistry* **287**, 8541-8551.
268. Zhang, J., and Corden, J.L. (1991). Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *The Journal of biological chemistry* **266**, 2290-2296.
269. Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S.L., Payne-Turner, D., Easton, J., Chen, X., Wang, J., Rusch, M., *et al.* (2012b). The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157-163.
270. Zhang, X., Wen, H., and Shi, X. (2012c). Lysine methylation: beyond histones. *Acta Biochim Biophys Sin (Shanghai)* **44**, 14-27.
271. Zhang, X., Yang, Z., Khan, S.I., Horton, J.R., Tamaru, H., Selker, E.U., and Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. *Molecular cell* **12**, 177-185.
272. Zheng, W., Ibanez, G., Wu, H., Blum, G., Zeng, H., Dong, A., Li, F., Hajian, T., Allali-Hassani, A., Amaya, M.F., *et al.* (2012). Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. *J Am Chem Soc* **134**, 18004-18014.
273. Zhu, X., He, F., Zeng, H., Ling, S., Chen, A., Wang, Y., Yan, X., Wei, W., Pang, Y., Cheng, H., *et al.* (2014). Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat Genet* **46**, 287-293.
274. Zlatanova, J., and Thakar, A. (2008). H2A.Z: view from the top. *Structure* **16**, 166-179.