EXAMINING THE BIOLOGICAL FUNCTIONS OF THE H3K36 METHYLATION STATES IN TRANSCRIPTION REGULATION AND THE POST-TRANSLATIONAL MODIFICATION LANDSCAPE OF SET2

Julia Veronica DiFiore

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology in the School of Medicine.

Chapel Hill
2020

Approved by:

Brian D. Strahl

Robert J. Duronio

Kerry S. Bloom

J. Mauro Calabrese

Ian J. Davis

ABSTRACT

Julia Veronica DiFiore: Examining the Biological Functions of the H3K36 Methylation States in Transcription Regulation and the Post-Translational Modification Landscape of Set2
(Under the direction of Brian D. Strahl)

To package DNA, eukaryotes fold and compact DNA into chromatin. The fundamental building block of chromatin is the nucleosome, which plays a crucial role in DNA accessibility. The unstructured N-terminal tails of histones can be post-translationally modified with chemical moieties such as acetylation and methylation, amongst others. These post-translational modifications (PTMs) can serve as binding sites for proteins that alter chromatin structure or affect the interactions between histones and DNA, thus making the DNA more or less accessible to other cellular machinery. Additionally, other proteins in the cell can be post-translationally modified with similar chemical moieties, which can alter their enzymatic activity, create binding sites for other proteins, or other outcomes. The role of PTMs in regulating cellular processes underscores their importance and the need to understand their function and regulation. One important chromatin-modifying enzyme in *Saccharomyces cerevisiae* is Set2, which is the sole enzyme that catalyzes mono-, di-, and trimethylation on lysine 36 on histone H3 (H3K36). Set2 and H3K36 methylation (H3K36me) are functionally important in transcription elongation, DNA damage repair, and mRNA spicing. However, most work examining Set2 and H3K36me was completed in wild-type cells or those completely lacking Set2 and H3K36me. Thus, our understanding of the functions of the different H3K36me states is incomplete. Similarly, while Set2-mediated methylation has been well studied, no studies have examined any PTMs on Set2 and their functional significance. Utilizing biochemical, genetic, and genomic assays, this work uncovers the unique and shared roles for the H3K36me states in a variety of cellular contexts. Importantly, H3K36me1/2 and H3K36me3 appear to have redundant roles in

repressing aberrant transcription during nutrient stress, thus providing flexibility during dynamic processes. Additional evidence demonstrates that Set2 is post-translationally modified and those PTMs have functional significance. Furthermore, the human homolog of Set2, SETD2, catalyzes H3K36me3 and robust evidence demonstrates that SETD2 is mutated in human cancers. Additional data presented in this work uncovers some of the structural and functional similarities and differences between Set2 and SETD2. Overall, the work presented here will further our understanding of Set2 and H3K36me in transcription, development, and disease.

To my family, whose sacrifices helped me get here and whose love kept me going, even in the darkest of times. This would not be possible without you.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Brian Strahl, for his guidance and support throughout the years. No matter the circumstances, Brian was always a source of enthusiasm. Whether it was troubleshooting experiments, analyzing new data, or discussing slightly crazy hypotheses, he could always be counted on for spirited discussion, new ideas, and a fresh outlook on my research. I would also like to thank my thesis committee members, Dr. Bob Duronio, Dr. Kerry Bloom, Dr. Ian Davis, and Dr. Mauro Calabrese. Initially, the thought of regularly presenting my work to a group of accomplished scientists was intimidating, but I quickly realized that they were all there with the intent to help my research be the best that it could be. I am grateful that our meetings were always full of thoughtful questions, ideas, and encouragement.

Throughout graduate school, I have been fortunate to work with many talented individuals. Deepak Jha took me under his wing when I first joined the lab and made it a fun place to be. Additionally, during my early years in graduate school, David Allison, Stephen McDaniel, Glenn Wozniak, Erin Shanle, and Jenny Kershner were always willing to give advice and fostered a supportive lab environment. Shelly Guo, Jake Bridgers, Alex Adams, Andrew Yoblinski, Brandon Boone, and the many other undergraduate students who have come through the lab have helped me become a better mentor and always provided a fresh perspective on things (lab and otherwise). Over the past year, I do not know where I would be without Kanishk Jain, Maddie Parker, Max Lowman, and Spencer Cooke. Our lunches together, dance parties, jokes, and nights on Franklin Street have kept me going when it seemed like my life would be a never-ending list of experiments and manuscript revisions. And finally, Hashem Meriesh and

Andrew Lerner have been in the lab with me since day one. We started off having no idea what we were doing and now, we sometimes know what we are doing. We did it, guys. We made it.

I have also had the opportunity to work with a lot of fantastic collaborators, both at UNC and otherwise. Catherine Fahey headed up a great group for a collaboration between the Strahl, Davis, and Rathmell labs. Working on that collaboration gave me confidence that I could meaningfully contribute to projects. Bing Li at Shanghai Jiao Tong University graciously provided some of the foundational data for my thesis work and has been helpful in providing additional data. Brenda Temple at the Structural Bioinformatics Core helped with protein modeling, which has proven to be instrumental in explaining the molecular basis underlying my thesis. Jeremy Simon and Travis Ptacek provided a thorough analysis of my RNA-seq data, which is another linchpin of my work. While normally rivals, collaborating with Dave MacAlpine, Heather MacAlpine, and Bonnie Chen at Duke has been great. Their expertise in MNase-seq has provided answers to long-standing questions and their enthusiasm makes for a fun collaboration. Finally, a recent project with Shu Zhang in the Dohlman lab has given new insight to some of the processes I have studied in graduate school. Shu's curiosity, hard work, and dedication make it easy to work with her and I look forward to hearing about her incredible discoveries in the years to come.

As Hermione Granger said in *Harry Potter and the Sorcerer's Stone*, "Books and cleverness! There are more important things - friendship and bravery." It is easy to loose sight of that in graduate school, but thankfully I found people who always made sure I remembered. Jen Kernan was my very first friend that I made in North Carolina and I am glad our first rotations and fall semester classes brought us together. Jess Simons has also always been there with an encouraging word, fun plans, crazy hospital stories, and her adorable pets. Without Jen and Jess, I would have had a lot less fun and not be the person and scientist I am today. I also absolutely would not have made it through graduate school without Austin Hepperla. He saved my ass on several occasions (figuratively and literally), can talk about anything for hours (in the

best way possible), and introduced me to *Game of Thrones*, *The Bachelor*, *How to Train Your Dragon*, and at least a dozen other things I now love. Michelle Engle has also been part of my graduate school tribe since the beginning and was always down for new adventures – whether locally or across the country. Her kindness and encouragement have gotten me through the many challenges of graduate school. The Yinzers – Casey Schmidt, Talia Hatkevich, Aaron Crain, and Chris Abdullah – watched many Steelers and Penguins games together and brought a small taste of home to North Carolina. Debbie and Ken Tunnel, Adam and Melanie Long, Holly Anderson, and many others at Chatham Community Church reminded me that I am so much more than what happens in lab. Most recently, Kevin Smith has been there to support, listen, and love me through all of this. He kept me fed, caffeinated, and laughing at his ridiculous dancing, puns, and unending episodes of *Brooklyn Nine-Nine*, and for that, I am forever grateful.

Finally, I would of course, not be here without my family. From my very first science fair in first grade when I was too scared to talk to anyone, to now, my parents have always supported my interest in math and science. And while my brother Matt was disappointed to find out that I was not curing cancer, he still enjoys hearing about my work and how the sausage gets made in academic research. They instilled the work ethic, perseverance, curiosity, and sense of humor that I would need to succeed. I am forever thankful for everything they have done for me and I hope to continue to make them proud.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

°	degree
3D	three dimensional
5-FOA	5-fluoroorotic acid
6-AU	6-azauracil
Å	angstrom
ac	acetyl
ALL	acute lymphocytic leukemia
AID	autoinhibitory domain
APC/C	anaphase promoter cyclosome complex
AWS	associated with SET
BLAST	basic local alignment search tool
BSA	Bovine serum albumin
C	celcius
CC	coiled coil
ccRCC	clear cell renal cell carcinoma
CD	circular dichroism
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation sequencing
CIS	cryptic initiation site
COSMIC	catalogue of somatic mutations in cancer
cryo-EM	cryogenic electron microscopy
CTD	C-terminal domain
CUT	cryptic unstable transcript
DNA	deoxyribonucleic acid

DRB	5,6-dichlorobenzimidazole 1-β-D-ribofuranoside
DSB	double strand break
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
ELM	eukaryotic linear motif
FACT	facilitates chromatin transcription complex
FAIRE-seq	formaldehyde assisted isolation of regulatory elements sequencing
FBS	fetal bovine serum
G1	Gap 1
G2	Gap 2
GST	glutathione S-transferase
Gy	gray
H2BK120	histone H2B lysine 120
H3K4	histone H3 lysine 4
H3K9	histone H3 lysine 9
H3K14	histone H3 lysine 14
H3K18	histone H3 lysine 18
H3K27	histone H3 lysine 27
H3K36	histone H3 lysine 36
H3K44	histone H3 lysine 44
H3K56	histone H3 lysine 56
H3K79	histone H3 lysine 79
H3S10	histone H3 serine 10
HA	hemagglutinin

HAT	histone acetyltransferase
HDAC	histone deacetylase complex
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMT	histone methyltransferase
hMSC	human mesenchymal stem cells
HR	homologous recombination
HRP	horseradish peroxidase
ICC	immunocytochemistry
IPTG	isopropyl β-d-1-thiogalactopyranoside
kDA	kilodalton
LANS	light-activated nuclear shuttle
LC-MS	liquid chromatography mass spectrometry
LEDGF	lens epithelium-derived growth factor
LON	long oligonucleotides
M	mitosis
MALDI-TOF	matrix assisted laser desorption/ionization-time of flight
MAP	mitogen-activated protein
MBF	*Mlu*I cell-cycle box binding factor
me	methylation
me1	mono-methylation
me2	di-methylation
me3	tri-methylation
mg	miligram
min	minute
mL	mililiter

mM	milimolar
MMR	mismatch repair
MNase-seq	microccocal nuclease sequencing
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MRX	Mre11, Rad50, Xrs2
MSL	male-specific lethal
MUD-PIT	multidimensional protein identification technology
NHEJ	Non-homologous end joining
NFR	Nucleosome free region
nM	nanomolar
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline-tween 20
PDB	protein data bank
PHD	plant homeodomain
Phe/Tyr	phenylalanine/tyrosine
PS	post-SET
PTM	post-translational modification
PVDF	polyvinylidene fluoride
PWWP	proline-tryptophan-tryptophan-proline domain
rep	replicate
RNA	ribonucleic acid

RNAPII	RNA polymerase II
RNA-seq	RNA sequencing
rpm	rotations per minute
rRNA	ribosomal RNA
RT-qPCR	quantitative reverse transcription polymerase chain reaction
S	DNA synthesis
SAH	S-adenosyl-L-homocycteine
SAM	S-adenosyl-L-methionine
SC	synthetic complete
SD	synthetic depleted
SDS-PAGE	sodium dodecyl sulface-polyacrylamide gel electrophoresis
SEM	standard error of the mean
SET	Su(var)3-9, Enhancer of zeste and Trithorax
SRAT	Set2-repressed antisense transcript
SRI	Set2 Rpb1 interacting domain
STAR	spliced transcripts alignment to a reference
SUMEB	SDS urea MOPS EDTA bromophenol blue
SUT	stable unannotated transcript
TALEN	TAL effector nucleases
TAP	tandem affinity purification
TFA	rribluroacetic acid
TOR1C	target of rapamycin 1 complex
tSETD2	truncated SETD2
ub	ubiquitination
µg	microgram

µl	microliter
µM	micromolar
UV	ultraviolet
V(D)J	variable diversity joining
qPCR	quantitative polymerase chain reaction
TCA	trichloroacetic acid
TEV	tobacco etch virus
T_m	melting temperature
WT	wild-type
XUT	Xrn1-sensitive unstable transcripts
YPD	yeast extract, peptone, dextrose

CHAPTER 1 – INTRODUCTION

The enormous diversity observed in the natural world underscores the complexity of life and has long pushed us to better understand its origins. Toward that end, in 1865, Gregor Mendel completed his studies on plant hybridization and decades later, the results would be recognized as foundational genetic principles (Mendel, 1901). Mendel's work and those who rediscovered it at the beginning of the 20th century founded the basis for understanding life, its diversity, and who we are at a molecular level (De Castro, 2016). Seminal work throughout the first half of the 20th century established a molecule called deoxyribonucleic acid (DNA) as the blueprint for all life on earth and the material that was responsible for passing traits from one generation to the next. Subsequent research established the "Central Dogma" of molecular biology, stating that DNA is transcribed into another molecule, ribonucleic acid (RNA), which is then translated into functional units called protein. However, recent discoveries have challenged this dogma and given us a more nuanced view of molecular biology and life itself (Gayon, 2016). Continued research will no doubt answer many of today's questions and raise more for future generations to grapple with.

The importance and complexity of DNA bring about several problems for all organisms. First, this immense amount of genetic information must be stored in an extremely small space: the nucleus of the cell. For humans, this entails storing over two meters of DNA into a nucleus that is ten μm in diameter. How is that level of compaction achieved? Molecular mechanisms must also exist to faithfully replicate DNA, repair any damage that occurs to it, as well as access DNA at the appropriate times for growth, repair, and other activities necessary for survival. And finally, all these processes must be coordinated so as to not interfere with each other. The work contained in this dissertation will focus on mechanisms for accessing DNA and how that

access is regulated through post-translational modifications (PTMs) on histone and non-histone proteins.

Histone Code Hypothesis

In order to package DNA, cells fold it into a higher order structure called chromatin. The basic unit of chromatin is the nucleosome, made up of two copies each of histones H2A, H2B, H3, and H4 with 147 base pairs of DNA wrapped around it (Luger et al., 1997). Histones can be post-translationally modified with chemical groups such as acetylation, methylation, phosphorylation, and ubiquitination, amongst others (Rothbart and Strahl, 2014). Many of these PTMs occur on the flexible histone tails, but some are found on the histone fold or globular domains (Cosgrove et al., 2004). These PTMs are functionally important in allowing or preventing cellular machinery to access DNA during cellular processes like transcription, replication, DNA damage repair, amongst others.

For years, it was thought that chromatin simply acted as a physical barrier to DNA-templated processes and organized DNA to fit into the nucleus. Beginning in the 1960s with the discovery of histone acetylation by Vincent Allfrey, evidence began to emerge that chemical modifications can affect DNA-histone contacts (Allfrey et al., 1964). For example, lysine acetylation changes histone tail from positively charged to neutral, therefore disrupted the binding with negatively charged DNA. Later, phosphorylation on serine 10 of histone H3 (H3S10) was shown to be important for two opposing functions: chromosome condensation during mitosis and chromatin decompaction in transcription (Bradbury, 1992; Mahadevan et al., 1991). This observation led to the proposal of the “histone code hypothesis,” which stated that chemical modifications on histones might form a code that direct specific functions on their own or in combination with other modifications (Strahl and Allis, 2000). In the last twenty years, numerous examples demonstrate the ability of histone modifications to alter chromatin structure and affect downstream biological outcomes.

Histone Lysine Methylation

One of the most well studied histone PTMs is lysine methylation. Since the detection of methylation on lysine 4 of histone H3 and the subsequent identification of SUV39H1 as the first lysine specific histone methyltransferase, methods have rapidly improved to detect histone lysine methylation, identify histone methyltransferases (HMT), and study their biological functions (Rea et al., 2000; Strahl et al., 1999). Histone lysine methylation is found on histone H3 at lysines 4, 9, 14, 18, 23, 27, 36, and 79, along with histone H4 at lysine 20 (Wozniak and Strahl, 2014). HMTs catalyze methylation at these residues by transferring a methyl group from the cofactor S-adenosyl-L-methionine (SAM) to the amino group of the lysine residue (Dillon et al., 2005). Each HMT has a specific lysine residue they target for methylation as well as specificity for how many methyl groups they add to the target lysine (mono-, di-, or trimethylation) (Black et al., 2012). Unlike acetylation, methylation does not change the charge of the lysine residue and is thought to primarily function as a way to recruit chromatin-modifying proteins (Beaver and Waters, 2016). By serving as a binding site for other proteins, histone lysine methylation has a diverse set of functions, particularly in gene transcription.

To precisely control gene transcription, different forms of histone lysine methylation are found in distinct regions of the genome (Weiner et al., 2015). Actively transcribed genes are marked by H3K4 methylation at their promoters, H3K36 methylation in the gene body, particularly at the 3' region, and H3K79 methylation throughout the gene body. In contrast, H3K9 methylation and H3K27 methylation are generally found at repressed genes (Wozniak and Strahl, 2014). The placement of these marks is controlled by the recruitment of their methyltransferase to distinct regions of the genome. Once the marks are catalyzed, reader proteins can bind and carry out their function. In addition to their precise localization, lysine methylation is removed to ensure proper genomic architecture is maintained (Gardner et al., 2011).

Set2 and H3K36 Methylation in Transcription

Regulation of Set2 and H3K36 Methylation

In *Saccharomyces cerevisiae* (hereafter, budding yeast), Set2 is the sole H3K36 methyltransferase and catalyzes mono-, di-, and trimethylation (H3K36me1/2/3) in actively transcribed genes (Strahl et al., 2002). In multicellular organisms, such as *Caenorhabditis elegans*, *Drosophila melanogaster*, and humans, there are multiple enzymes that catalyze H3K36me. In *D. melanogaster* and humans, certain enzymes are specific for H3K36me1 and H3K36me2, while the Set2 homologs in those organisms catalyze only H3K36me3 *in vivo* (Venkatesh and Workman, 2013). Nevertheless, budding yeast serves as an excellent model system for Set2 and H3K36me since the loss (*set2*Δ) or mutation of one enzyme can alter H3K36me globally and the consequent changes can be studied.

Set2 catalyzes H3K36me in actively transcribed genes and is targeted there in part through its interaction with the phosphorylated serine 2 and serine 5 on the C-terminal domain (CTD) of RNA polymerase II (RNAPII) (Li et al., 2003; Xiao et al., 2003; Youdell et al., 2008). While serine 5 is phosphorylated at the promoter by Kin28, serine 2 phosphorylation depends on Ctk1 and Bur1 (Zaborowska et al., 2016). Loss of Ctk1 results in reduced Set2 protein levels and loss of H3K36me (Chu et al., 2006; Fuchs et al., 2012; Youdell et al., 2008). Similarly, bur1 mutants have reduced H3K36me3 (Chu et al., 2006). Another protein that is important for Set2 regulation is the histone chaperone Spt6, which also associates with the phosphorylated CTD of RNAPII (Bortvin and Winston, 1996; Diebold et al., 2010; Sun et al., 2010). Critically, the association of Spt6 with the CTD stabilizes Ctk1 protein levels and Ctk1 and serine 2 phosphorylation also maintain Spt6 protein levels (Dronamraju and Strahl, 2014). Moreover, in *spt6-1004* mutants, there is no detectable Set2 or H3K36me (Youdell et al., 2008).

In addition to regulation by transcriptional machinery, contacts on the nucleosome itself affect Set2 and H3K36me. A recent cryo-EM structure of Set2 in Chaetomium thermophilum shows extensive contacts with H3 αN, H3 tail, and H2A C-terminal tail in order to position Set2

over H3K36. Furthermore, the associated with SET (AWS) domain (discussed below) interacts with H2BK120ub, which increases Set2 activity on nucleosomes (Bilokapic and Halic, 2019). Overall, a network of transcriptional machinery and key nucleosomal contacts are important for targeting Set2 and H3K36me to the correct genomic locations, maintaining normal protein levels, and positioning the enzyme over its substrate.

Set2 Protein Structure

Set2 contains several, well-defined protein domains (Figure 1.1). Two of the most crucial domains to Set2 function are the catalytic SET domain and the Set2 Rpb1 Interacting (SRI) domain. The catalytic SET domain is responsible for methylating H3K36. It is located at the N-terminus and is flanked by the associated with SET (AWS) domain and the post-SET (PS) domains (Strahl et al., 2002). Both the AWS and PS domains contain three conserved cysteines that coordinate zinc near the active site and are necessary for catalysis (Cheng et al., 2005). The SRI domain is located at the C-terminus of Set2. It is necessary for binding to the phosphorylated serine 2 and serine 5 residues on the CTD of RNAPII during transcriptional elongation (Kizer et al., 2005). Without the SRI domain, Set2 cannot catalyze H3K36me3 and only low levels of H3K36me2 are present (McDaniel et al., 2017). Interestingly, in the absence of the SRI domain, Set2 is still recruited to gene bodies, indicating that other mechanisms work along with the SRI domain to recruit Set2 to genes (Youdell et al., 2008).

Several other domains also contribute to Set2 function. The H4 interacting domain, located at the extreme N-terminus, is important for nucleosomal binding, specifically contacting H4K44. In the absence of H4 binding, Set2 cannot catalyze H3K36me2 or H3K36me3, likely due to improper positioning and decreased residence time over the nucleosome (Du et al., 2008). In the middle of Set2, there is the auto-inhibitory domain (AID). The AID antagonizes the SET and SRI domains to control the catalytic activity of Set2 and ensure H3K36me is correctly localized in gene bodies (Wang et al., 2015). Lastly, Set2 contains two protein-protein binding

domains, the WW domain and the coiled-coil (CC) domain. Both are located between the AID and the SRI domain, however, there are no known functions for either of these domains in budding yeast. The WW domain is named for two conserved tryptophans (W) and tends to bind to proline-rich regions in other proteins (Sudol et al., 1995). Thus, the WW domain could mediate non-histone interactions of Set2. CC domains are often involved with subunit oligomerization and phosphorylation within the domain can modulate its binding ability (Burkhard et al., 2001). Consequently, the CC domain in Set2 could be functionally relevant for oligomerizing Set2 and dynamically regulated by phosphorylation.

H3K36 Methylation Interacting Proteins

H3K36 methylation acts as a binding site for several chromatin-modifying enzymes and prevents the binding of a histone chaperone (Figure 1.2). The known proteins that bind to H3K36me bind through either a PWWP or chromo domain, which are common reader domains of lysine methylation (Rona et al., 2016; Yap and Zhou, 2011; Yun et al., 2011). Overall, the combined function of the H3K36 effector proteins is to create a hypoacetylated environment after RNAPII-mediated transcription and prevent aberrant transcriptional events (Venkatesh and Workman, 2013).

One protein that cannot bind in the presence of H3K36me is Asf1, a histone chaperone that contributes to the deposition of H3K56ac by Rtt109 (Tsubota et al., 2007). H3K56ac is associated with histone exchange, particularly over promoters (Williams et al., 2008). However, Asf1 cannot bind to peptides that are di- or trimethylated at H3K36, which likely contributes to suppressing histone exchange and maintaining a hypoacetylated chromatin environment in the wake of RNAPII (Venkatesh et al., 2012).

A crucial protein that binds to H3K36me is Eaf3 and it is part of the Rpd3S histone deacetylases complex (HDAC), which also includes Rpd3, Rco1, Sin3, and Ume1 (Carrozza et al., 2005). Eaf3 binds to H3K36me2 and H3K36me3 through its chromodomain and

allosterically activates Rpd3 to remove H3 and H4 acetylation (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Merker et al., 2008; Ruan et al., 2015). Additionally, binding by the PHD of Rco1 to unmodified H3 helps target Rpd3S to gene bodies (Li et al., 2007; McDaniel et al., 2016). Overall, the Rpd3S complex creates a hypoacetylated environment to prevent aberrant transcription from initiating within gene bodies.

Another important protein that can bind H3K36me is Ioc4, which is part of the Isw1b ATP-dependent chromatin-remodeling complex (Vary et al., 2003). Through its PWWP domain, Ioc4 binds to H3K36me3 *in vivo* and H3K36me3 and H3K36me2 *in vitro*. By associating with H3K36me, Ioc4 localizes Isw1b to sites of active transcription to reposition nucleosomes (Maltby et al., 2012; Smolle et al., 2012). Isw1b nucleosome remodeling activity helps position nucleosomes to be an appropriate substrate for Rpd3S HDAC activity and create a hypoacetylated chromatin environment (Lee et al., 2013).

The most recently discovered protein that binds to H3K36me is Pdp3 and it is part of the NuA3 HAT complex, which also includes Sas3, Yng1, Taf14, Eaf6, and Nto1 (Gilbert et al., 2014). Pdp3 binds to H3K36me3 through its PWWP domain and targets NuA3 to gene bodies (Gilbert et al., 2014; Martin et al., 2017). It is intriguing that H3K36me recruits both a HDAC (Rpd3S) and HAT (NuA4) complex to actively transcribed genes. One possibility is that the HDAC and HAT work in context-dependent manners. NuA3 is targeted to sites with both H3K36me3 and H3K4me1/2/3 through the PWWP domain of Pdp3 and the PHD of Yng1, respectively (Martin et al., 2017). Therefore, sites with both H3K36me3 and H3K4me1/2/3 would be more prone to HAT activity, while HDAC activity would be more prevalent at sites lacking the combinatorial marking preferred by NuA3. Overall, the precise regulation of histone PTMs through combinatorial histone marks supports the idea of the histone code hypothesis.

Cryptic Transcription

A well-documented function for Set2 and H3K36 methylation is to prevent cryptic transcription. When normal chromatin architecture is disrupted, transcription can initiate from intragenic regions instead of canonical promoter regions (Kaplan et al., 2003). Cryptic transcripts typically arise from nucleosome free regions (NFRs) and can be transcribed in the sense or antisense direction (Neil et al., 2009; Xu et al., 2009). Several classes of cryptic transcripts have been defined, including cryptic unstable transcripts (CUTs), stable unannotated transcripts (SUTs), and Xrn1-sensitive unstable transcripts (XUTs). CUTs are rapidly degraded by the exosome or cytoplasmic decay pathways. CUTs were first detected in exosome mutants (*rrp6*Δ) and later detected in 5'-to-3' exonuclease mutants (*xrn1*Δ) and cytoplasmic decapping mutants (*dcp1*Δ or *dcp2*Δ) (Thompson and Parker, 2007; Wyers et al., 2005). Later, a specific class of transcripts degraded by Xrn1 was defined as XUTs (van Dijk et al., 2011). The rapid detection and degradation of these transcripts suggest they may have a negative effect on cellular processes. In contrast, SUTs are more resistant to decay mechanisms and are detectable in wild-type cells (Xu et al., 2009). Overall, cryptic transcripts are widespread in the budding yeast genome and differ in their sensitivity to degradation pathways.

In addition to the rapid degradation of cryptic transcripts, there are several mechanisms to suppress their transcription. In particular, Set2 and H3K36me recruit the Rpd3S HDAC to create a hypoacetylated environment in the wake of RNAPII and prevent additional polymerases and transcription factors from binding and initiating transcription from intragenic regions (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2007). Importantly, the nucleosomes must be properly spaced for Rpd3S to bind and activate its deacetylase activity. To accomplish this, Isw1b binds to H3K36me and remodels the nucleosomes so they are properly spaced for Rpd3S and the complex can bind to its preferred substrate (Lee et al., 2013). Additionally, H3K36me prevents the histone chaperone Asf1 from binding and thereby suppresses Rtt109-dependent H3K56ac and histone exchange (Venkatesh et al., 2012). These

mechanisms have been mostly studied in *set2*Δ cells and the contributions of individual H3K36me states is still not well understood (Figure 1.3).

While Set2 and H3K36me play a critical role in preventing cryptic transcription, there are other important factors that repress cryptic transcription. Spt6 is a histone chaperone that binds H3/H4 dimers and requires H2A/H2B-binding FACT to reassemble nucleosomes after transcription by RNAPII (Bortvin and Winston, 1996; Mccullough et al., 2015). Cryptic transcripts were first identified in an *spt6-1004* mutant and subsequently observed in other mutant strains, including *SPT16*, a member of the FACT complex. *spt6-1004* strains showed increased MNase sensitivity compared to wild-type, indicating a depletion of histones, altered chromatin structure, or both (Kaplan et al., 2003). Future work revealed that spt6 mutants had decreased nucleosome occupancy at the 5' ends of genes and increased cryptic transcription from these regions compared to wild-type (Dronamraju et al., 2018). FACT and Spt6 also work together to prevent the promoter-specific histone variant H2A.Z from being incorporated into gene bodies and promoting cryptic transcription (Jeronimo et al., 2015). In addition to Spt6 and FACT complex members, a genome-wide screen identified 50 genes that were important for repressing cryptic transcription. Most of these genes included histones, chromatin modifying proteins, and transcription elongation factors (Cheung et al., 2008). Robust evidence demonstrates that chromatin structure is a key regulatory step in repressing cryptic transcription and its misregulation contributes to aberrant transcription.

While improving sequencing technologies have allowed for the detection of cryptic transcripts and a mechanistic understanding of how they arise, the function of cryptic transcripts remains unclear (Churchman and Weissman, 2011; Nielsen et al., 2019). Recently, a group of transcripts that is suppressed exclusively by Set2 and H3K36 methylation were identified and defined as SRATs. These transcripts were detected in wild-type cells, but they were more abundant in *set2*Δ cells; however, their function was not identified (Venkatesh et al., 2016). Another study showed that during carbon source shifting, the Set2 and Rpd3S pathway is

important for repressing transcription at certain genes. Interestingly, most genes regulated by the Set2 and Rpd3S pathway have an overlapping lncRNA over their promoter. The transcription of the lncRNA brings Set2, H3K36me, and Rpd3S to the gene and contributes to its repression (Kim et al., 2016). Similarly, work from our lab found that under nutrient deprivation conditions, *set2*Δ cells produced cryptic transcripts whose presence was correlated with decreased sense transcription at corresponding genes (McDaniel et al., 2017). In accordance with a decrease in sense transcription, others detected decreased protein abundance for genes whose promoters overlap with a SUT (Huber et al., 2016). Such evidence suggests that cryptic transcription may function in the cellular response to stress and regulate transcript and protein abundance, but more work is needed to better understand the relationship between cryptic transcription and stress response. Interestingly, some cryptic transcripts appear to be translated, however the function of these truncated polypeptides is still not understood (Cheung et al., 2008; Wei et al., 2019). Overall, cryptic transcripts can expand the yeast transcriptome and proteome and future work focusing on their functionality will be key in understanding the significance of this phenomenon.

Additional Functions in Budding Yeast and Set2 Homologs

Additional Functions in Budding Yeast

In addition to transcription elongation, Set2 and H3K36me have functions in DNA damage repair, the cell cycle, aging, and mRNA splicing. Set2 and H3K36me function in the early stages of checkpoint activation and require the interaction with RNAPII to function at sites of DNA damage (Jha and Strahl, 2014; Winsor et al., 2013). In the absence of Set2 and H3K36me, cells are sensitive to phleomycin, a double strand break (DSB) inducing agent, and also show reduced γ-H2A.X, increased H4 acetylation, and increased H3 and Htz1 retention at DSB sites. The altered chromatin structure found at DSB sites makes *set2*Δ cells more prone to homologous recombination (HR) than non-homologous end joining (NHEJ). (Jha and Strahl,

2014). Future work examining the possibility that H3K36me recruits specific repair machinery will further our understanding of Set2 and H3K36me in DNA damage response.

During the cell cycle, Set2 and H3K36me function to repress cryptic transcription at cell cycle genes. Repressing cryptic transcription and thereby preventing transcriptional interference ensures that the cell cycle genes are properly expressed. However, in *set2*Δ and K36A strains, cells show delayed progression through G1 and rapidly progress through S phase. Interestingly, Set2 and H3K36me3 are most abundant during G2/M and targeted for destruction during M phase, with evidence suggesting that the APC/C^{CDC20} complex degrades Set2 (Dronamraju et al., 2017). Similar to its function in repressing cryptic transcription during carbon source shifting and nutrient stress, Set2 and H3K36me also contribute to maintaining the precise transcriptional programming of the cell cycle (Dronamraju et al., 2017; Kim et al., 2016; McDaniel et al., 2017).

Global chromatin alterations have been linked to aging, including changes to H3K36me. An overarching model is that reduced nucleosome occupancy contributes to expression of genes that are normally repressed (Hu et al., 2014). Interestingly, cells with H3K36 mutants (H3K36R or H3K36E) or *set2*Δ displayed a decreased life span, while cells lacking the H3K36 demthylase, Rph1, had an increased life span. The decreased longevity in H3K36me deficient and *set2*Δ cells was attributed to the increased acetylation across gene bodies, thus more open chromatin environment, and increased cryptic transcription found in those cells (Sen et al., 2015). Overall, the role of Set2 and H3K36me in maintaining a proper chromatin environment and repressing cryptic transcription serves as an important mechanism in aging.

When looking at splicing defects in *set2*Δ cells, there is decreased splicing efficiency, which relies on the interaction between Set2 and RNAPII (Leung et al., 2019; Sorenson et al., 2016). Additionally, some evidence suggests that Eaf3 bound to H3K36me recruits splicing factors Prp45 and Prp19 to actively transcribed genes (Leung et al., 2019). While less than 5% of genes have introns in budding yeast, those genes account for 30% of transcripts (Ares et al., 1999). Since H3K36me occurs co-transcriptionally, its involvement in splicing is not surprising,

and better mechanistic insight into how it interacts with and recruits the spliceosome will improve our understanding of the combined roles of chromatin, transcription, and splicing in gene expression.

Schizosaccharomyces pombe

Schizosaccharomyces pombe, or fission yeast, is a distant cousin to budding yeast. It is considered a more ancient yeast species since fewer evolutionary changes occurred in fission yeast than budding yeast since diverging from their common ancestor. Overall, more genes are conserved between fission yeast and mammals than between budding yeast and mammals. The similarities and differences between budding yeast and fission yeast provide valuable evolutionary insight. Conserved mechanisms between the two species indicate a mechanism that is likely conserved across eukaryotes, while differences between the two suggest more diversity in other eukarytoes (Hoffman et al., 2015). In fission yeast, as in budding yeast, Set2 is the sole enzyme responsible for H3K36me1/2/3 and binds to phosphorylated serine 2 on the RNAPII CTD through its SRI domain (Kizer et al., 2005; Strahl et al., 2002; Suzuki et al., 2016). Likewise, Set2 and H3K36me influence DSB repair choice in fission yeast, tipping the balance in favor of NHEJ. However, data in fission yeast suggests that H3K36me peaks in S/G2 of the cell cycle and Set2 protein levels are constant throughout (Pai et al., 2014). The difference in cell cycle regulation is intriguing, especially considering the human homolog, SETD2, is regulated similarly to budding yeast Set2 (discussed below) (Dronamraju et al., 2017). Interestingly, Set2 and H3K36me were implicated in the DNA replication checkpoint pathway, but recent data indicates that cells lacking Set2 and H3K36me are capable of normal DNA replication checkpoint signaling (Kim et al., 2008; Pai et al., 2017). Instead, H3K36me in fission yeast seems to be important for MBF complex binding, which controls the expression of genes important for the G1/S transition (Pai et al., 2017). Overall, the similarities and differences

between fission and budding yeast continue to better our understanding of the molecular mechanisms regulated by Set2 and H3K36me.

Caenorhabditis elegans

In *Caenorhabditis elegans*, there is no longer a single enzyme that catalyzes H3K36me, but two: MET-1 and MES-4. MET-1 catalyzes H3K36me co-transcriptionally, while MES-4 does not require an association with RNAPII and is important for germline development (Andersen and Horvitz, 2007; Bender et al., 2006; Furuhashi et al., 2010; Rechtsteiner et al., 2010). Previously, it was thought that only MET-1 was capable of generating H3K36me3, but recent evidence demonstrates that MES-4 can also catalyze H3K36me3. MES-4 is the major H3K36me3 methyltransferase in early embryos and requires the H3K36me3 transmitted to the progeny by the sperm and oocyte for its recruitment, thus establishing a mechanism for the epigenetic memory of H3K36me3 across generations (Kreher et al., 2018). Genome-wide analysis shows that H3K36me3 is enriched over exons, a pattern that is similar to humans (discussed below) (Kolasinska-Zwierz et al., 2009). Additionally, *C. elegans* is one of the premiere model systems for studying development and aging because of their short life span and ability to produce many progeny (Corsi et al., 2015). Multiple studies observed shorter lifespans in *C. elegans* that have decreased H3K36me3 levels (Pu et al., 2015; Sen et al., 2015). Mechanistically, it appears that H3K36me3 acts to prevent cryptic transcription and maintain proper gene expression, similar to budding yeast (Sen et al., 2015). Continued work in *C. elegans* will help elucidate the function of H3K36me in multicellular organisms, particularly in splicing, aging, and epigenetic memory.

Drosophila melanogaster

Drosophila melanogaster has long been a powerful model organism due to their easily manipulated genomes and short life span, making them particularly compelling model systems

for development and cell biology (Hales et al., 2015). In *Drosophila*, like in *C. elegans*, there are two enzymes that catalyze H3K36me. Mes-4 catalyzes H3K36me1 and H3K36me2, while dSet2, the homolog of budding yeast Set2, catalyzes H3K36me3 (Venkatesh and Workman, 2013). dSet2 and H3K36me3 is required for proper development and like in budding yeast, it associates with the hyperphosphorylated CTD of RNAPII (Bell et al., 2007; Stabell et al., 2007). In H3K36R mutants, there is an increase in H4 acetylation, but no cryptic transcripts detected, indicating differences in how cryptic transcription in metazoans is regulated compared to budding yeast. Interestingly, evidence suggests that H3K36me is important post-transcriptionally and functions in proper mRNA maturation (Meers et al., 2017). Another function of H3K36me in Drosophila that is distinct from its function in budding yeast is in recruiting the MSL complex to the male X chromosome for dosage compensation (Larschan et al., 2007). A recently developed histone gene replacement platform in Drosophila is an excellent system to study the function of specific histone residues and will help continue to increase our knowledge of H3K36me in metazoans (McKay et al., 2015).

Humans and Cancer Relevance

In humans, there are several enzymes that catalyze H3K36me, underscoring the increased complexity of gene regulation in multicellular metazoans. NSD1, NSD2, NSD3, SETMAR, SMYD2, and ASH1L catalyze H3K36me1 and H3K36me2, while SETD2 catalyzes H3K36me3 (McDaniel and Strahl, 2017). SETD2 is the human homolog of yeast Set2 and was initially identified in a yeast two-hybrid screen for interacting partners of huntingtin protein (Faber et al., 1998; Sun et al., 2005). Like other Set2 homologs, SETD2 interacts with the hyperphosphorylated CTD of RNAPII and is important during development (Hu et al., 2010; Sun et al., 2005). Since its identification, SETD2 has been linked to many cellular processes, such as splicing, DNA repair, and cell cycle regulation, amongst others.

Given the well-established role for H3K36me3 in transcription, it is not surprising that H3K36me3 is also important for splicing. The first evidence of its role in splicing came from examining alternative splicing of the FGFR2 gene in human cells. Exon IIIb of FGFR2 is included in PNT2 cells and exon IIIc is excluded, while the opposite occurs in hMSCs; exon IIIb is excluded and exon IIIc is included. The PTB protein regulates the alternative splicing of FGFR2 by binding to silencing elements around exon IIIb and ensuring its exclusion. Interestingly, H3K36me3 was enriched at the FGFR2 gene in hMSCs, cells where exon IIIb is excluded. When SETD2 was overexpressed and H3K36me3 levels increased, there was increased exon IIIb exclusion, establishing a causal relationship between H3K36me3 and alternative splicing (Luco et al., 2010). Additionally, in tumor cells with mutated SETD2 and reduced H3K36me3, there was increased intron retention (Simon et al., 2014). The exact mechanism for H3K36me3 in splicing is not known. H3K36me3 may recruit MRG15, the human homolog of Eaf3, to splice sites which may in turn recruit other proteins for splicing (Luco et al., 2010). Tumor cells with mutated SETD2 had changes in chromatin accessibility in addition to splicing defects, suggesting the ability of H3K36me3 to recruit chromatin remodelers may also contribute to its role in splicing (Simon et al., 2014).

Several studies have demonstrated a function for SETD2 and H3K36me3 in DNA damage repair pathways and the cell cycle. During mismatch repair (MMR), the mismatch recognition protein hMutSα binds to H3K36me3 through its PWWP domain. Cells with decreased SETD2 compared to wild-type have microsatellite instability, which is common in certain forms of cancers (Li et al., 2013). Additionally, H3K36me3 can recruit LEDGF, CtIP, and RAD51, which are critical for end resection during in HR (Carvalho et al., 2014; Kanu et al., 2015; Pfister et al., 2014). The involvement of SETD2 with HR is intriguing, considering the role of budding yeast Set2 in promoting NHEJ over HR (Jha and Strahl, 2014). However, in the cell cycle, SETD2 exhibits a conserved function with budding yeast Set2. SETD2 protein levels are highest in G2/M and lowest in G1 and S phase. Additionally, SETD2 is important for

transcriptional regulation of cell cycle genes and targeted by APC/C for destruction (Dronamraju et al., 2017).

In addition to its functions in splicing, DNA repair, and the cell cycle, SETD2 has been implicated in a variety of other cellular processes. Both de novo DNA methyltransferases, DNMT3a and DNMT3b, bind H3K36me3 through their PWWP domains (Dhayalan et al., 2010; Rondelet et al., 2016). In vitro work and studies in mouse stem cells demonstrate that DNMT3a and DMNT3b activity is increased by binding to H3K36me3 and targets the enzymes to gene bodies (Baubec et al., 2015; Dhayalan et al., 2010; Morselli et al., 2015). Interestingly, methylation by DNMT3b represses cryptic transcription and depends upon recruitment and activation by H3K36me3 (Neri et al., 2017). Another silencing mechanism that SETD2 is functionally related to is PRC2-mediated gene repression. The Tudor domain of PCL1 can bind H3K36me3 and recruit PRC2, the enzyme responsible for H3K27me3 and transcriptional repression at many developmental loci (Cai et al., 2013). Recent work in mice has also shown that SETD2 and H3K36me3 are necessary for lymphocyte development and V(D)J recombination (Ji et al., 2019). Finally, in addition to its role in methylating H3K36, SETD2 has been connected to tubulin methylation and cytoskeleton reorganization (Park et al., 2016). All of these functions underscore the importance of SETD2 and H3K36me3 in maintaining genomic integrity.

With SETD2 and H3K36me3 playing an integral role in many cellular functions, it is no surprise that SETD2 is loss or mutated in many forms of cancer. SETD2 mutations in cancer were first observed in clear cell renal cell carcinoma (ccRCC) (Dalgliesh et al., 2010). SETD2 is located on chromosome 3p and one allele of chromosome 3p is frequently loss in ccRCC. Along with SETD2, VHL, PBRM1, and BAP1 are located on chromosome 3p and the combined chromosomal loss, inactivation of VHL, and mutation in a chromatin-modifying enzyme characterizes many ccRCC tumors (Duns et al., 2012; Gerlinger et al., 2012; The Cancer Genome Atlas, 2013; Zbar et al., 1987). In addition to ccRCC, SETD2 mutations are found in a

myriad of cancers, including lung adenocarcinoma, endometriod carcinoma, and hematopoietic cancers, amongst others (Jiang et al., 2018; Li et al., 2016; Yuan et al., 2017; Zhu et al., 2014). In addition to the loss or mutation of SETD2, mutations to H3K36 have recently emerged in several types of cancer. Initially, a lysine to methionine (H3K36M) mutation was identified and characterized in chondroblastomas, a type of bone cancer. Tumors with H3K36M showed a global decrease in H3K36me2 and H3K36me3 along with a redistribution of H3K27me3 and PRC1/PRC2-mediated silencing. While only 1 allele out of the 32 encoding for H3K36 was mutated to H3K36M, the mutation acted as a dominant-negative by trapping SETD2 and inhibiting its methyltransferase activity (Fang et al., 2016; Lu et al., 2016). Based on the crystal structure of the SETD2 catalytic domain bound to H3K36M peptide, conformational changes in the PS domain contribute to the increased association with H3K36M and decreased enzymatic activity (Yang et al., 2016; Zhang et al., 2017). In addition to chondroblastoma, H3K36M mutations are observed in pediatric soft tissue sarcoma and head and neck squamous cell carcinoma (Lu et al., 2016; Papillon-Cavanagh et al., 2017). Continued efforts to understand how H3K36M mutations contribute to cancer development will likely reveal similar mechanisms as cancers with SETD2 mutations.

Despite the lack of mechanistic understanding of how SETD2 and H3K36m3 contribute to tumorigenesis, their functions in transcription, DNA repair, and the cell cycle have already been implicated. Approximately 25% of genes expressed in one cohort of ccRCC tumors had RNA processing defects (Simon et al., 2014). Transcriptional data from The Cancer Genome Atlas demonstrates that transcriptional read-through (transcription occurring past the 3' termination site) is common in tumors with SETD2 mutations and correlates with worse patient outcomes. Read-through transcription can interfere with transcription of the neighboring gene and create RNA chimeras, thus interfering with normal gene expression (Grosso et al., 2015). H3K36M mutations redistribute PRC1 and suppresses gene expression (Lu et al., 2016). At sites of DNA damage, ccRCC and ALL cells failed to properly load repair machinery and lose

important checkpoint activation signaling (Carvalho et al., 2014; Kanu et al., 2015; Mar et al., 2017). Additionally, cell cycle genes are improperly expressed when SETD2 and H3K36me3 are absent in a variety of tumors (Dominguez et al., 2016). Examining the interplay of transcription, DNA repair, and cell cycle regulation by SETD2 and H3K36me3 will be critical in understanding their contributions to cancer development.

Non-histone PTMs and their Biological Functions

PTMs on non-histone proteins have important biological functions. Cellular processes such as transcription, the cell cycle, and DNA repair rely on non-histone PTMs for proper activation, recruitment, and destruction of key proteins involved (Biggar et al., 2017; Duan and Walther, 2015; Narita et al., 2019). As discussed above, phosphorylation on the CTD of RNAPII regulates the recruitment of transcriptional machinery. Without proper CTD phosphorylation, key proteins are not recruited and normal gene expression cannot occur (Zaborowska et al., 2016). Interestingly, advances in mass spectrometry have revealed that chromatin-modifying enzymes are also post-translationally modified (Biggar et al., 2017). Furthermore, changes to the PTMs on chromatin-modifying enzymes have already been associated with disease (Habibian and Ferguson, 2019).

Three well-studied PTMs on non-histone proteins include methylation, acetylation, and phosphorylation. Methylation and acetylation are found on lysine and arginine residues, while phosphorylation occurs on serine, threonine, and tyrosine residues (Buuh et al., 2017). The PTMs contribute to their enzymatic activity, protein stability, and subcellular localization, amongst other functions (Biggar and Li, 2015; Narita et al., 2019; Ubersax and Ferrell, 2007). Several examples have recently emerged, such as Gcn4 acetylation of Rsc4 preventing Rsc4 from binding to its normal substrate and carrying out its remodeling function to phosphorylation of NuA4 and its role in DNA damage repair (Cheng et al., 2018; VanDemark et al., 2007). The

interplay between non-histone PTMs, transcription, and other cellular processes is a key regulatory feature across organisms.

Concluding Remarks and Contributions of This Work

Since the proposal of the Histone Code Hypothesis, significant progress has been made toward understanding how histone PTMs function in a variety of cellular processes. Furthermore, studies focused on non-histone PTMs have also increased our understanding of cell signaling and gene expression regulation. A well-studied methyltransferase, Set2, and its catalytic product, H3K36me, are conserved from yeast to humans, thus underscoring their importance in transcriptional regulation and other cellular processes. While substantial progress has been made in understanding the functions of Set2 and H3K36me, various questions remain unanswered. Many studies of Set2 and H3K36me take place in the context of wild-type or *set2*Δ cells; however there has not been a comprehensive examination of the different H3K36me states and their functions. In Chapter 2, I describe biochemical, genetic, and genomic assays that elucidate the unique and shared functions of the H3K36me states. Furthermore, no studies have examined the PTMs that occur on Set2 and how those affect Set2 function. In Chapter 3, I present biochemical and phenotypic data that provides evidence for PTMs on Set2 and their functions. As discussed earlier in this Chapter, the human homolog of Set2, SETD2, is commonly mutated in cancer, particularly ccRCC. In Appendix A, I present data highlighting the similarities and differences between Set2 and SETD2 in their structure and function. Finally, Chapter 4 discusses the contributions of this work in the larger context of transcriptional regulation and future experiments that will increase our understanding of Set2 and H3K36me. Overall, the work of this dissertation provides new insight into the regulation and function of Set2 and H3K36me.

Figures

Figure 1.1: Set2 Domain Map

The H4 interacting domain is at the N-terminus (green). The catalytic domain is comprised of the Associated with SET domain (AWS, yellow), the SET domain (tan), and the Post-SET domain (PS, orange). In the middle of the protein is the autoinhibitory domain (blue). In the C-terminus, there are three protein-protein interacting domains, including the WW domain (red), the coiled-coiled domain (teal), and the Set2 Rpb1 Interacting domain (SRI, purple).

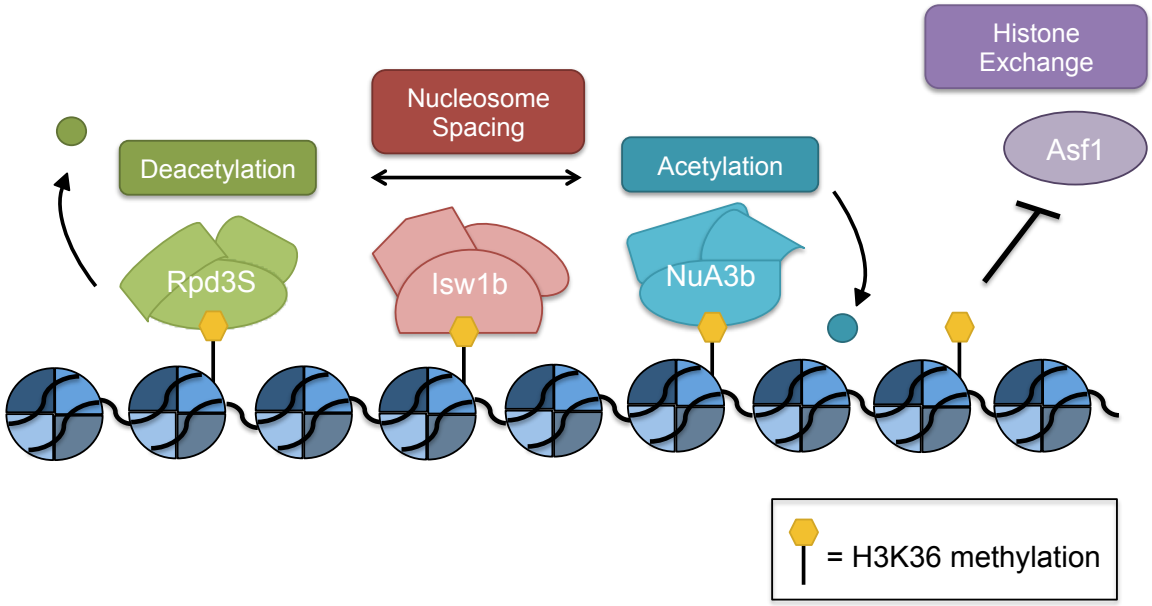


Figure 1.2: H3K36 Methylation Interacting Proteins

Rpd3S (green), Isw1b (pink), and NuA3b (blue) complexes bind to H3K36me (yellow) and carry out their respective functions on nucleosomes. Asf1 (purple) cannot bind to nucleosomes marked with H3K36me.

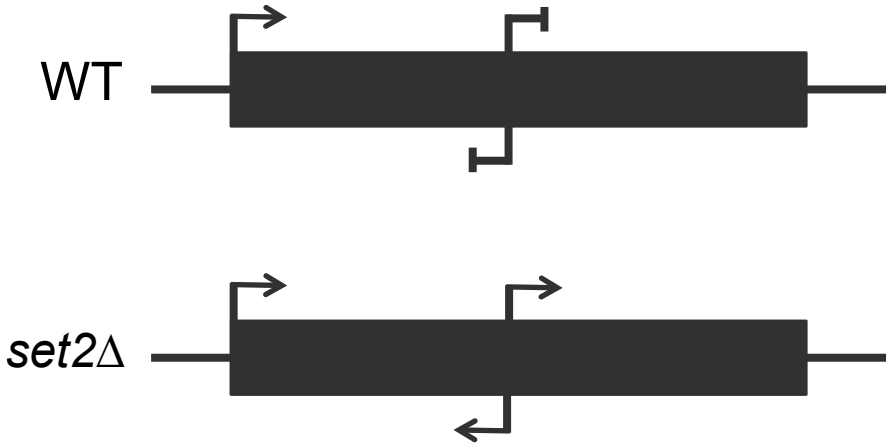


Figure 1.3: Schematic of Cryptic Transcription in wild-type and *set2*Δ

In wild-type cells, transcription starts from the canonical 5' promoter and is repressed from within gene bodies. In *set2*Δ cells, cryptic transcription is observed and can occur bidirectionally from within the gene body.

CHAPTER 2 – UNIQUE AND SHARED ROLES FOR HISTONE H3K36 METHYLATION IN TRANSCRIPTION REGULATION FUNCTIONS

Summary

Set2 co-transcriptionally methylates lysine 36 of histone H3 (H3K36), producing mono-, di-, and trimethylation (H3K36me1/2/3). These modifications recruit or repel chromatin effector proteins important for transcriptional fidelity, mRNA splicing, and DNA repair. However, it was not known whether the different methylation states of H3K36 have distinct biological functions. We used engineered forms of Set2 that produced different methyl states to identify unique and shared functions for H3K36 modifications. Although H3K36me1/2 and H3K36me3 were functionally redundant in many *SET2* deletion phenotypes, H3K36me3 had a unique function related to Bur1 kinase activity and FACT complex function. Further, during nutrient stress, either H3K36me1/2 or H3K36me3 repressed high levels of histone acetylation and cryptic transcription that arises from within genes. Our findings uncover the potential for regulation of diverse chromatin functions by different H3K36 methylation states.

Introduction

Histone post-translational modifications affect a great variety of DNA-templated processes. Methylation, acetylation, and other modifications are added to histones by chromatin-modifying enzymes (Rothbart and Strahl, 2014; Soshnev et al., 2016). These chemical modifications can alter histone-DNA contacts and they often serve as docking sites for effector proteins. The coordinated effort of histone modifications, chromatin-modifying enzymes, and effector proteins helps RNA polymerase II (RNAPII) to gain access to DNA for transcription,

replication, and DNA repair (Papamichos-Chronakis and Peterson, 2013; Smolle and Workman, 2013). A similar coordinated effort is employed to make the genome unreachable by RNAPII at appropriate times to prevent aberrant transcription.

Set2 is a chromatin-modifying enzyme that contributes to prevention of inappropriate transcription. Set2 methylates histone H3 at lysine 36 (H3K36) (McDaniel and Strahl, 2017; Strahl et al., 2002; Venkatesh and Workman, 2013). In *Saccharomyces cerevisiae* (hereafter, budding yeast), Set2 is responsible for all forms of H3K36 methylation, mono-, di-, and trimethylation. Set2 binds to the C-terminal domain (CTD) of transcribing RNAPII and catalyzes H3K36 methylation (H3K36me) in actively transcribed genes (Kizer et al., 2005; Xiao et al., 2003). H3K36me provides docking sites for several proteins, such as Rpd3S, a histone deacetylase complex, and Isw1b, a nucleosome remodeler. Rpd3S is recruited to chromatin by the PHD fingers in Rco1 and the activity of Rpd3S is stimulated by binding of the Eaf3 chromodomain to H3K36me2 (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2007; McDaniel et al., 2016; Ruan et al., 2015). Additionally, Isw1b associates with chromatin by way of the Ioc4 PWWP domain binding to H3K36me3 (Maltby et al., 2012; Smolle et al., 2012). Collectively, these processes ensure transcriptional fidelity by preventing transcription initiation from within gene bodies, a process known as cryptic transcription.

In the absence of Set2 and H3K36me, both sense and antisense cryptic transcription occur across the genome. Cryptic transcription tends to be a consequence of bi-directional transcriptional events at cryptic promoters within gene bodies (Carrozza et al., 2005; Churchman and Weissman, 2011; Joshi and Struhl, 2005; Lickwar et al., 2009; Neil et al., 2009; Xu et al., 2009). Precisely how cryptic sites become accessible in *SET2* deletion mutants (*set2*Δ) is not fully understood; however, a lack of Set2-mediated H3K36me results in increased histone exchange along with increased histone acetylation – both phenomena that may promote recruitment of RNAPII initiation factors and chromatin remodelers such as RSC that promote chromatin disruption (Li et al., 2007; Pattenden et al., 2010; Smolle et al., 2012; Venkatesh et

al., 2012). From a functional standpoint, cryptic transcription in the absence of Set2 does not result in significant growth defects when examined in nutrient-rich medium, although cell cycle progression is partially disrupted (Dronamraju et al., 2018; Kim et al., 2016; Lenstra et al., 2011; McDaniel et al., 2017; Venkatesh et al., 2016). In contrast, during nutrient deprivation or carbon source shifting, cryptic antisense transcription can occur at a level sufficient to impair normal transcription and lead to improper gene expression and poor cell growth (Kim et al., 2016; McDaniel et al., 2017).

Although the functions of Set2 and H3K36 methylation have been studied intensely, there are no studies that have thoroughly interrogated which functions of Set2 are directed by the different (mono- di-, and tri-) H3K36 methylation states. Previous reports suggest that H3K36me3 is dispensable for suppression of cryptic transcription (Hacker et al., 2016; Li et al., 2009; Youdell et al., 2008). In contrast, *set2* mutants that harbor only H3K36me1 exhibit cryptic transcription, suggesting that the main functions of H3K36 methylation occur through H3K36me2 (Hacker et al., 2016). However, it is not known whether H3K36me3 function is unique or overlapping with the other H3K36me states.

In this study, we engineered the SET domain of Set2 so that it performed only H3K36me1, H3K36me1/2, or H3K36me3, thus affording a unique opportunity to interrogate the functions of different methylation states. We found that Set2 that generated H3K36me3 and H3K36me1/2 can act redundantly to rescue several canonical phenotypes associated with set2Δ, e.g., caffeine sensitivity and DNA damage response. In contrast, Set2 that generated only low levels of H3K36me1 did not rescue *set2*Δ phenotypes. However, only H3K36me3 rescued other *set2*Δ phenotypes, e.g., *bur1*Δ or *spt16-11* bypass. Intriguingly, strains that produced either H3K36me1/2 or H3K36me3 could largely suppress cryptic transcription during nutrient stress. Approximately 60% of the identified cryptic transcript initiation sites contained a degenerate TATA box motif, a fact that suggested how those sites become susceptible to cryptic transcription in the absence of H3K36me. Additionally, H3K36me1/2 or H3K36me3 could

function to ensure proper levels of H3K27ac and H3K56ac in genes, pointing to a potential mechanism for preventing inappropriate transcriptional initiation by control of nucleosome remodeling and histone exchange. In sum, our data provide key evidence for the independent and overlapping functions of H3K36me1/2 and H3K36me3 in chromatin biology and in transcriptional regulation.

Results

Phe/Tyr Switch in Set2 separates H3K36 Methylation States *in vitro*

Although previous studies have assessed the effects of losing or limiting H3K36 methylation, no study has comprehensively examined the individual functions of each H3K36 methylation state produced by Set2. To create a system in which we could control production of H3K36 methylation states, we employed a structural mutagenesis approach for Set2 based on the Phe/Tyr switch phenomenon that alters the SET domain lysine-binding pocket (Cheng et al., 2005; Collins et al., 2005). Briefly, two well-positioned aromatic residues within the SET domain (a phenylalanine [Phe, F] and a tyrosine [Tyr, Y]) can be switched to the opposite aromatic residue to control the size of the catalytic domain. An F to Y mutation creates steric hindrance in the active site with lysine trimethylation, thereby limiting the enzyme to only mono- and dimethylation. However, a Y to F change creates more space and promotes the catalysis of trimethylation. In Set2, Y149 and F234 were predicted to comprise the Phe/Tyr switch (Cheng et al., 2005; Collins et al., 2005) (Figure 2.1A).

To determine whether Set2 Y149 and F234 can function as a Phe/Tyr switch, we created a structural model of the Set2 SET domain by using MODELLER and the crystal structure of SETD2 (PDB ID 5JLB) as the template (Webb and Sali, 2016; Yang et al., 2016). The template and alignment were the top hit generated by I-TASSER (Zhang, 2008). The model confirmed that Y149 and F234 are part of the lysine-binding pocket (Figure 1B, left panel) and

configured to enable Set2 to mono-, di-, and trimethylate H3K36. We then investigated this model for the effect of deploying the Phe/Tyr switch. Modeling showed that the Y149F mutation opened the SET domain binding pocket and eliminated hydrogen bonding between the hydroxyl group of tyrosine and H3K36me (Figure 2.1B, middle panel). These changes would permit rapid rotation of the target lysine in the presence of S-adenosylmethionine (SAM) and likely increase the potential for higher methylation states on H3K36. Conversely, the F234Y substitution created steric hindrance between the tyrosine hydroxyl group and the target lysine, thus suggesting that this mutant would limit H3K36me3 (Figure 2.1B, right panel). These findings are consistent with these residues encompassing a working Phe/Tyr switch in Set2.

Lastly, we measured the effect of the Phe/Tyr switch on Set2 enzymatic activity. We performed histone methyltransferase (HMT) activity assays with recombinant Set2 (wild-type, *set2-Y149F*, and *set2-F234Y*), SAM, and recombinant Xenopus nucleosomes; methylation states were detected by Western blot analysis with antibodies specific to the three modifications (Figures 2.1C and 2.1D). We found that *set2-Y149F* catalyzed predominately H3K36me3, whereas *set2-F234Y* catalyzed H3K36me1 and H3K36me2 (Figure 2.1D). Overall, these results demonstrated that the SET domain of Set2 contains a functional Phe/Tyr switch that can be used to produce specific H3K36 methylation states.

Phe/Tyr Switch Mutations in Set2 are Tools to Separate H3K36 Methylation States *in vivo*

To test the effect of the Set2 Phe/Tyr switch on H3K36 methylation *in vivo*, we generated yeast strains harboring endogenous *set2-Y149F* or *set2-F234Y* mutations. For controls, we included a *set2*Δ strain and a catalytic domain mutant strain (*set2-H199L*) that was capable of producing only low levels of H3K36me1 *in vivo* (Hacker et al., 2016; Jha and Strahl, 2014). The *set2-Y149F* mutation predominately resulted in H3K36me3 and a low level of H3K36me1 (Figure 2A). In contrast, the *set2-F234Y* mutation limited Set2 methylation to H3K36me1/2

(Figure 2.2A). As previously described, the absence of Set2 resulted in complete loss of H3K36 methylation and the catalytic domain mutant, *set2-H199L*, showed only a low level of H3K36me1 (Figure 2.2A and Figure 2.3A). The SET domain mutations did not alter the amount of Set2 protein (Figure 2.2B), however, global H3K36 methylation differed from wild-type. The *set2-Y149F* mutant had an average of 56% of the wild-type H3K36me3 (Figure 2.2C). *set2-F234Y* showed an H3K36me2 signal similar to wild-type (93%), but a 5-fold increase in H3K36me1 (Figures 2.2D and 2.2E). *set2-H199L* had a reduced H3K36me1 signal compared with wild-type. Overall, these results confirmed that the Phe/Tyr switch in Set2 functioned *in vivo* and, although global H3K36me levels varied from wild-type, we could largely control different methylation states.

Because some of our downstream genetic analyses would require expression of *SET2* from plasmids, we also examined the effect of the *set2* mutants in two exogenous expression systems: expression of *SET2* and the *set2* mutants from their native promoter and overexpression of the same genes from the *ADH1* promoter. Western blot analyses for H3K36 methylation in these two systems revealed two important finings. First, the *set2-Y149F* and *set2-F234Y* mutants, regardless of expression system, recapitulated the initial *in vivo* findings that the Phe/Tyr switch could control the methylation states of H3K36 (Figure 2.3B). Second, and surprisingly, overexpression of *SET2* via the *ADH1* promoter resulted in robust overexpression of Set2 protein compared with expression from its endogenous promoter, yet global H3K36 methylation levels in the wild-type or *set2* mutants were similar to or lower than normal. Thus, although it was possible to greatly overexpress Set2, the amount of methylation that Set2 can produce on chromatin must have a limit, perhaps due to a finite number of Set2 binding sites on the CTD of RNAPII, restricted histone accessibility, limited SAM availability, or a combination of these factors.

Overall, these results were consistent with the *in vitro* data and they demonstrated that the Phe/Tyr switch mutants were specific for performing predominately H3K36me3 or

H3K36me1/2. These provided tools to examine *in vivo* the functions of different methylation states of H3K36.

Distinct Methylated Forms of H3K36 are Deposited within or near Transcribed Regions of Genes

We next asked whether the *set2* mutants still deposited H3K36me within or near the coding regions of genes in a manner consistent with its known localization and function (Pokholok et al., 2005; Rao et al., 2005; Weiner et al., 2015). We first confirmed by co-immunoprecipitation that the *set2* mutants still associated normally with phosphorylated RNAPII (Figure 2.3C). Consistent with RNAPII association, ChIP-qPCR of all H3K36me forms in the *set2* mutants confirmed that they were deposited within or near the coding regions of several loci tested (*STE11*, *PMA1*, and *TDH3,* see Figures 2.4A, 2.4E, and 2.4I). Additionally, the H3K36me states detected by ChIP-qPCR were consistent with the aforesaid *in vitro* HMT and Western blot assays (Figures 2.1C and 2.2A).

Closer comparison of the differently methylated H3K36 forms with wild-type revealed several key observations. As shown in Figures 2.4B, 2.4F, and 2.4J, there were no significant differences in the H3K36me3 levels between wild-type and *set2-Y149F* at all loci tested. However, for H3K36me2, we observed differences between wild-type and *set2-F234Y* at *STE11* and *PMA1*. A significantly higher H3K36me2 level was detected in the middle of *STE11* in *set2-F234Y* (Figure 2.4C), whereas, at *PMA1*, a significantly higher H3K36me2 level was detected across the entire gene body (Figure 2.4G). In contrast, we did not detect any significant difference in H3K36me2 levels between *set2-F234Y* and wild-type at *TDH3* (Figure 2.4K). As expected from the high H3K36me1 Western blot signal, *set2-F234Y* had significantly higher levels of H3K36me1 in the 3' end of *STE11* and across the entirety of *PMA1* and *TDH3* compared with wild-type (Figures 2.4D, 2.4H, and 2.4L). Finally, although global levels of H3K36me1 were lower in *set2-H199L*, we observed nearly wild-type levels of H3K36me1 in

set2-H199L at the loci examined. At *STE1,* we did not find significant differences in H3K36me1 between *set2-H199L* and wild-type (Figure 2.4D and Figure 2.5). However, at *PMA1*, there was significantly more H3K36me1 deposited by *set2-H199L* in the gene body (Figure 2.4H Figure 2.5D). In contrast, we detected less H3K36me1 at the 5' end of *TDH3* in *set2-H199L* compared with wild-type, although the rest of that locus did not show a significant difference between the mutant and wild-type. Overall, these results established that the levels of H3K36me in the *set2* mutants differed from wild-type at certain loci, but the results further confirmed the specificity of the *set2* mutants and indicated that differentially methylated forms of H3K36 were localized within or near the bodies of transcribed genes.

H3K36me1/2 and H3K36me3 Function Redundantly in Some Cellular Contexts

To determine whether the different forms of H3K36me are responsible for distinct biological outcomes, we examined the sensitivity of set2 methylation mutants to drugs that affect *set2*Δ cell growth. We tested caffeine (a TOR1C and MAP kinase inhibitor), rapamycin (a TOR1C inhibitor), phleomycin (a double-strand break-inducing agent), and 6-azauracil (6-AU, a transcriptional elongation inhibitor) (Jha and Strahl, 2014; Kizer et al., 2005; McDaniel et al., 2017). Like previous work, we found that *set2*Δ and *set2-H199L* mutant cells were sensitive to caffeine, rapamycin, and phleomycin, and they were resistant to 6-AU (Figures 2.6A and 2.6B and Table 2.1). Intriguingly, similar to wild-type, the *set2-Y149F* and *set2-F234Y* mutants rescued the drug-induced growth defects from caffeine, rapamycin, and phleomycin (Figure 2.6A and Table 2.1). However, only *set2-Y149F* was as sensitive to 6AU as wild-type, with *set2-F234Y* having an intermediary phenotype between wild-type and *set2*Δ (Figure 2.6B). Thus, in some cellular contexts, such as TORC1 signaling and double-strand break repair, H3K36me3 and H3K36me1/2 were functionally redundant, and a low level of H3K36me1 alone could not suffice.

Genetic Interactions of *SET2* with *BUR1* and *SPT16* Reveal Unique Functions for H3K36me1/2 and H3K36me3

In addition to phenotypes of *set2*Δ cells challenged with drugs, deletion of *SET2* bypasses the lethal or slow growth phenotypes associated with the inactivation or deletion of members of the BUR kinase and FACT complexes (Biswas et al., 2006; Chu et al., 2006; Keogh et al., 2003). Both complexes regulate aspects of transcription; the BUR kinase complex promotes RNAPII elongation and the FACT complex is responsible for nucleosome disruption and reassembly. Cells lacking *BUR1* (*bur1*Δ) are not viable, but they survive when *SET2* is also deleted (*bur1*Δ *set2*Δ). Likewise, cells harboring the FACT mutant allele *spt16-11* grow extremely slow at 34°C, and deletion of Set2 bypasses this phenotype. Therefore, we examined our set2 methylation mutants to determine whether the different H3K36me states would bypass the *bur1*Δ and *spt16-11* phenotypes. We confirmed that the presence of *set2*Δ in the *bur1*Δ or *spt16-11* cells rescued the lethal or slow growth phenotypes, whereas exogenously expressed wild-type *SET2* did not rescue growth (Figures 2.6C and 2.6D and Table 2.1). Surprisingly, cells containing the *set2-Y149F* allele that catalyzed only H3K36me3 phenocopied wild-type Set2, whereas the *set2-F234Y* and *set2-H199L* alleles that catalyzed H3K36me1/2 or low levels of H3K36me1, respectively, phenocopied the *set2*Δ cells (Figures 2.6C and 2.6D and Table 2.1). Thus, these findings indicated that H3K36me3 has a function distinct from H3K36me2 and H3K36me1 in relation to Bur1 kinase activity and FACT complex function.

H3K36me1/2 and H3K36me3 Have Unique and Shared Functions in Repressing Cryptic Transcription at Reporter Loci

An extensively documented phenotype of *set2*Δ cells is the appearance of sense and antisense cryptic transcripts that arise from intragenic regions (McDaniel and Strahl, 2017). Although our studies and others have revealed that H3K36me3 is dispensable for preventing cryptic transcription (Hacker et al., 2016; Li et al., 2009; Youdell et al., 2008), it was not known whether H3K36me3 alone can support suppression of cryptic transcription in the absence of

H3K36me1/2. To answer this question, we used two reporter genes, *FLO8* and *STE11,* that each have internal cryptic initiation sites (CIS) and the *HIS3* gene integrated out-of-frame and downstream of the CIS. When the *FLO8* and *STE11* loci are transcribed from canonical promoters, *HIS3* is not transcribed in-frame, and the cells cannot survive on media lacking histidine (Silva et al., 2012; Wang et al., 2015). However, when cryptic transcription occurs at the CIS, *HIS3* is transcribed in-frame, and cells survive without histidine (Figure 2.6E and 2.6F). Thus, we first confirmed that cells with wild-type *SET2* suppressed cryptic transcription and did not grow on media lacking histidine, whereas *set2*Δ cells did not suppress cryptic transcription at either locus (Figures 2.6G and 2.6H and Table 1). At the *FLO8* locus, which was under the GAL promoter and constitutively expressed in cells growing on galactose-containing media, both H3K36me3 and H3K36me1/2 were sufficient to suppress cryptic transcription similar to wild-type *SET2* (Figure 2.6G and Table 2.1). Surprisingly, at the *STE11* locus, expressed from its native promoter and at a lower level than *FLO8*, only H3K36me3 prevented cryptic transcription, whereas cells with H3K36me1/2 or a low level of H3K36me1 did not suppress (Figure 2.6H and Table 2.1). Thus, at *STE11,* H3K36me3 alone is sufficient to suppress cryptic transcription. Additionally, these results suggested that H3K36me3 and H3K36me1/2 have different, but sometimes overlapping, functions in preventing cryptic transcription. Particularly, highly and lowly expressed genes may have different requirements for H3K36me.

During Nutrient Deprivation, H3K36me1/2 or H3K36me3 Prevent Antisense Transcription

Having established that different forms of H3K36me are necessary to repress cryptic transcription at specific loci, we examined this fact globally. Previous work from our lab demonstrated that robust bidirectional cryptic transcription occurs at 439 genes in *set2*Δ cells during nutrient deprivation (McDaniel et al., 2017). We used this system as a tool to learn the genome-wide requirement of the different H3K36me states to prevent cryptic transcription. We

performed stranded RNA-seq at 0 and 60 minutes following nutrient deprivation in wild-type, *set2*Δ, *set2-Y149F*, *set2-F234Y*, and *set2-H199L* cells. We examined sense and antisense transcriptional signals in the genomic regions that surround the high- and intermediate-confidence cryptic initiation sites of the aforesaid 439 genes. After 60 minutes of nutrient deprivation, we observed a global increase in sense and antisense transcription surrounding the CIS in *set2*Δ and *set2-H199L* cells (Figure 2.7A). Interestingly, cryptic transcription was not completely repressed in *set2-Y149F* or *set2-F234Y*; both mutants showed a slight increase in antisense transcription upstream of the CIS compared with wild-type (Figure 2.7A). When we separated the CIS-containing genes based on gene expression levels, we observed that *set2-Y149F* or *set2-F234Y* repressed sense and antisense cryptic transcription similar to wild-type in highly and lowly expressed genes (Figure 2.8A). In contrast, *set2*Δ and *set2-H199L* did not repress any form of cryptic transcription in highly or lowly expressed genes (Figure 2.8A). Importantly, we verified that nutrient deprivation did not alter existence of the H3K36 methylation states catalyzed by the *set2* mutants (Figure 2.7A). These results revealed that H3K36me1/2 or H3K36me3 can function to prevent cryptic transcription genome-wide and low levels of H3K36me1 do not have a large part in repressing cryptic transcription.

Next, we wanted to further characterize the antisense transcription that occurred in the *set2* mutants. We previously showed that, in *set2*Δ, 92 of 121 high-confidence CISs exhibited robust antisense transcription that extended to the transcription start-site. Here, we found that, relative to wild-type, *set2*Δ and *set2-H199L* cells exhibited increased antisense signal that extended to the transcription start site, irrespective of gene length, whereas this was rarely observed in *set2-Y149F* and *set2-F234Y* cells or when comparing wild-type replicates (Figure 2.7B and Figure 2.8B). Previously, we had suggested that, by extending into gene promoter regions, this antisense transcription interfered with and reduced sense transcription (McDaniel et al., 2017). To examine this idea for the *set2* mutants, we compared the sense and antisense transcriptional signals between wild-type and each of the *set2* mutants, focusing specifically on

the regions between the CIS and transcription start sites. Any signal in these regions should reflect normal sense transcription and cryptic antisense transcription, but not cryptic sense transcription because cryptic sense transcription would only occur downstream of the CIS. Indeed, we frequently observed decreases in sense transcription and concomitant increases in antisense transcription, with the strongest effects on both strands occurring in *set2*Δ and *set2-H199L* cells (Figure 2.7C). The effects observed here were more pronounced than the effects we observed between two wild-type replicates compared as a control (Figure 2.8C). Moreover, many of the same genes exhibited similar effects for all the *set2* mutants, particularly, *set2*Δ, *set2-H199L*, and *set2-F234Y* cells (Figure 2.7D). Together, these data suggested that, during nutrient deprivation, H3K36me1/2 or H3K36me3 suppresses cryptic antisense transcription, which appears to drive the downregulation of sense transcription from certain CIS-containing genes.

Lastly, we wondered whether the sequences surrounding the CIS harbored any particular sequence motifs that may explain why cryptic transcription initiates there and not elsewhere. We searched the sequences ±100 bp of the 439 high- and intermediate-confidence CIS and found a significant enrichment of a (T/C)AAT motif, which may represent a degenerate TATA box element (Figure 2.7E) (Basehoar et al., 2004; Lubliner et al., 2013). We did not observe this motif for a random sampling of 439 low-confidence CIS (Figure 2.8D). Therefore, these data suggested that cryptic transcription may initiate from TATA-like elements, but their suppression requires H3K36me1/2 or H3K36me3. Mechanistically, H3K36me1/2 and H3K36me3 may be capable of recruiting HDACs and preventing histone exchange to maintain a repressive chromatin environment near these TATA-like elements.

Alteration of Global H3K27ac and H3K56ac in *set2* Methylation Mutants

Given the ability of H3K36me1/2 and H3K36me3 to suppress antisense transcription during nutrient stress, we next sought to determine whether these methylation states would also suppress the co-transcriptional histone acetylation associated with the appearance of cryptic transcripts (Carrozza et al., 2005; Du and Briggs, 2010; Joshi and Struhl, 2005). H3K36 methylation limits co-transcriptional histone acetylation in two main ways, by recruitment and activation of the Rpd3S deacetylation complex that removes H3 and H4 N-terminal tail acetylation (e.g., H3K27ac) and by the ability of H3K36me to repel the histone exchange factor Asf1, which helps deposit the histone exchange-associated mark H3K56ac (Merker et al., 2008; Venkatesh et al., 2012; Williams et al., 2008). Before evaluating the effects of the *set2* mutants on histone acetylation, we first examined whether nutrient stress affected the amounts of Set2 protein or degrees of H3K36 methylation. The *set2* mutants showed only a slight reduction in Set2 protein levels after nutrient stress, but they maintained their H3K36 methylation specificity (Figure 2.9A and Figures 2.10A-D). Interestingly, global H3K36me1 levels were about equal under normal growth conditions and nutrient stress, whereas H3K36me2 and H3K36me3 increased after nutrient stress in *set2-F234Y* and *set2-Y149F*, respectively.

Examination of global H3K27ac levels in wild-type and our *set2* mutants revealed several interesting observations. Under normal growth conditions, H3K27ac levels were higher in *set2*Δ and the *set2* mutants compared to wild-type (Figures 2.9A and 2.9B). Global H3K27ac levels remained steady in wild-type cells after nutrient stress and H3K27ac level in *set2-Y149F*, *set2-F234Y* , and *set2-H199L* decreased to that of wild-type. In contrast, H3K27ac levels remained high in *set2*Δ strains after nutrient depletion. These results were consistent with the known function of Set2-dependent H3K36 methylation in recruiting and stimulating the activity of the Rpd3S histone deacetylase complex.

When looking at H3K56ac, most strains had similar levels in normal growth conditions, with *set2-F234Y* showing a slightly lower level (Figures 2.9A and 2.9C). Intriguingly, wild-type

and *set2-Y149F* showed close to a 50% reduction in H3K56ac after nutrient stress. However, *set2*∆, *set2-F234Y*, and *set2-H199L* did not show a similar reduction in H3K56ac. Thus, the data suggests that during nutrient stress, H3K36me3 functions to reduce global H3K56ac levels because only strains containing H3K36me3 had reduced H3K56ac.

H3K36me1/2 and H3K36me3 are Important for Proper H3K27ac and H3K56ac Localization

After investigating the global effects on H3K27ac and H3K56ac levels during nutrient stress, we next sought to determine how the localization of these histone acetylation events would be affected in wild-type and our *set2* mutants. Using ChIP-qPCR, we examined H3K27ac and H3K56ac at *STE11* and *SPB4* in wild-type, *set2*∆, *set2-Y149F*, *set2-F234Y* , and *set2-H199L* strains before and after nutrient stress. Both loci showed nutrient stress-induced cryptic transcription in our RNA-seq data set and both genes produced cryptic transcripts as assayed by Northern blot (Cheung et al., 2008; McDaniel et al., 2017). On the basis of the lengths of cryptic transcripts detected in the RNA-seq data, we designed qPCR primers to cover the predicted CIS (for *STE11,* primer set 4; for *SPB4*, primer set 5) and other locations across the genes (Figures 2.11A and 2.11F and Figures 2.12A and 2.12F).

When examining H3K27ac localization, we found several differences between the *set2* mutants. In *set2*∆, there were higher H3K27ac levels compared to wild-type near the CIS of *STE11* and *SPB4* in the 3' regions of *STE11*, particularly after nutrient stress (Figures 2.11B and 2.11G). These data were consistent with previous reports that, in the absence of H3K36 methylation, Rpd3S was unable to deacetylate nucleosomes (Drouin et al., 2010; Govind et al., 2010). In contrast, for *set2-Y149F* and *set2-F234Y*, changes in H3K27ac were different at *STE11* and *SPB4*. At *STE11* after nutrient stress, both *set2-Y149F* and *set2-F234Y* had significantly higher H3K27ac compared with wild-type near the CIS and immediately 3' (Figures 2.11C and 2.11D). However, at *SPB4*, H3K27ac levels in both mutants were nearly identical to

wild-type before and after nutrient stress (Figures 2.11H and 2.11I). One commonality in *set2-Y149F* and *set2-F234Y* was the significantly lower H3K27ac levels compared to wild-type at the promoter before nutrient stress. Similarly, *set2-H199L* had significantly reduced H3K27ac at the promoter regions of *STE11* and *SPB4* before nutrient stress (Figure 2.11E and 2.11J). However, the changes in H3K27ac near the promoter may be due to Rpd3L and Set3C recruitment, as those complexes localize to 5' ends of genes (Kim et al., 2012; Shi et al., 2007; Wang et al., 2011). Nonetheless, in *set2-H199L*, H3K27ac levels were higher near the CIS of *STE11* and 3' regions of both genes, particularly in *STE11.* These results demonstrated that, in certain contexts, H3K36me1/2 or H3K36me3 can ensure proper H3K27ac levels in gene bodies to repress cryptic transcription.

The *set2* mutants also showed clear differences in H3K56ac localization before and after nutrient stress. In *set2*Δ, there were significantly higher H3K56ac levels compared with wild-type near the CIS and in the 3' regions of *STE11* and *SPB4*, particularly after nutrient stress (Figures 2.12B and 2.12G). Interestingly, we also found higher levels of H3K56ac near the CIS of *STE11* under normal growth conditions. This observation suggested that, in the absence of H3K36 methylation, sites near TATA-like elements are prone to histone exchange, and this exchange is exacerbated during nutrient stress. In contrast, *set2-Y149F* and *set2-F234Y* generally showed similar levels of H3K56ac compared to wild-type near the CIS and in the 3' regions of *STE11* and *SPB4* (Figures 2.12C, 2.12D, 2.12H, and 2.12I). However, there were significant differences in each mutant in the upstream noncoding, promoter, and promoter proximal regions; both *set2-Y149F* and *set2-F234Y* had decreased H3K56ac signals compared with wild-type. Although *set2-H199L* showed robust cryptic transcription, there were no significant differences in H3K56ac compared to wild-type near the CIS or in the 3' coding regions, suggesting that alteration of something other than histone exchange was responsible for the appearance of cryptic transcription at these loci during nutrient stress (Figures 2.12E and 2.12J). Taken

together, these data supported a model in which H3K36me1/2 or H3K36me3 can ensure proper H3K56ac levels in gene bodies and repress histone exchange.

Discussion

During transcriptional elongation, Set2 catalyzes H3K36me1, H3K36me2, and H3K36me3 and suppresses cryptic transcription (Kim et al., 2016; McDaniel and Strahl, 2017; Venkatesh et al., 2016). However, it was unclear whether the three methylation states of H3K36 have distinct biological functions and which methyl state(s) is important for suppression of cryptic transcription. We have demonstrated that H3K36me1/2 and H3K36me3 have unique and shared functions, particularly regarding suppression of antisense transcription. By taking advantage of a functional Phe/Tyr switch in the SET domain of Set2, we controlled the methylation products generated by this enzyme and established the *in vivo* functions of different H3K36 methylation states. We observed that in, certain cellular contexts, such as rescuing the *set2*Δ phenotype caused by absence of the Bur1 kinase or Spt16 of the FACT complex, H3K36me1/2 and H3K36me3 had distinct activities. However, in other contexts, such as sensitivity to caffeine, rapamycin, and phleomycin, H3K36me1/2 and H3K36me3 were functionally equivalent. Similarly, in nutrient stress, H3K36me1/2 or H3K36me3 suppressed antisense cryptic transcription from within gene bodies. Without H3K36me1/2 or H3K36me3, we observed antisense cryptic transcription across the genome, and many loci also had decreases in sense transcription, in agreement with previous findings (McDaniel et al., 2017). We also found evidence of a degenerate TATA box motif in loci with CIS, suggesting that DNA sequence contributes to the sensitivity of these sites to cryptic transcription. Mechanistically, H3K36me1/2 and H3K36me3 have an important function in ensuring proper levels of H3K27ac and H3K56ac in genes, particularly during nutrient stress. By maintaining proper histone acetylation levels within genes, cells prevent inappropriate nucleosome remodeling and histone exchange that

likely governs cryptic initiation events (Pattenden et al., 2010). Overall, these findings reveal different functions of the H3K36 methylation states in a diverse set of transcriptional programs.

Our study shows that, in certain cellular contexts, H3K36me1/2 and H3K36me3 are interchangeable, thus providing flexibility to dynamic processes. This equivalency may exist partly because Eaf3 and Ioc4 bind to both H3K36me2 and H3K36me3, as previously shown by *in vitro* assays (Carrozza et al., 2005; Joshi and Struhl, 2005; Li et al., 2009; Maltby et al., 2012). Additionally, both H3K36me2 and H3K36me3 peptides are refractory to Asf1 binding (Venkatesh et al., 2012). The recruitment or repelling of these effector proteins by sites with either H3K36me2 or H3K36me3 maintains a proper transcriptional response in certain contexts, even in the absence of one of those marks. Although previous data supported a model wherein H3K36me3 was dispensable for certain cellular processes, particularly suppression of cryptic transcription, it was counterintuitive that the cell would use energy to catalyze an unnecessary mark (Hacker et al., 2016; Li et al., 2009; Youdell et al., 2008). Our findings suggest that, instead of H3K36me3 being dispensable, it often has a redundant function with H3K36me1/2, thus affording the cell flexibility, particularly in response to stress.

However, our study provides evidence that, in certain contexts, H3K36me3 has a unique function for which the other H3K36me forms cannot substitute. Particularly, H3K36me1/2, but not H3K36me3, genetically interacted with Bur1 and Spt16. With Bur1 acting upstream of Set2 during transcription, it is unclear how Bur1 activity relates to specific H3K36me states. For the FACT complex, *in vitro* results indicated that Spt16 can bind H3K36me2 peptides, but not H3K36me3, peptides (Venkatesh et al., 2012). Perhaps nucleosomes with H3K36me3 are refractory to FACT binding, which may have protected wild-type and *set2-Y149F* cells from FACT-dependent nucleosome remodeling. Further work to ascertain how H3K36me3 functions in the Bur1 pathway and in FACT complex-mediated nucleosome exchange will further highlight how specific histone modifications facilitate distinct processes, as proposed by the histone code hypothesis (Strahl and Allis, 2000).

We were unable to ascribe a function for H3K36me1 in any of the Set2 phenotypes that we examined. This lack of functionality may have been because the low levels of H3K36me1 observed in *set2-H199L* were not enough to elicit an activity or there simply was no function for H3K36me1 in these phenotypes; we cannot distinguish between these possibilities. Notably, all known reader domains of H3K36me interact with H3K36me2 and H3K36me3, not H3K36me1 (McDaniel and Strahl, 2017). Pryde et al. (2009) suggested that H3K36me1 recruits the replication factor Cdc45 to origins. We did not interrogate this particular activity, thus, we cannot rule out that H3K36me1 may function in replication. H3K36me1 may be targeted by a yet-to-be identified reader that functions in origin activity.

Interestingly, we identified an enriched sequence motif that resembles a degenerate TATA box, within 100 bp of almost 60% of high- and intermediate-scoring CIS. Although these sites were identified previously, it was not known what predisposed them to cryptic initiation events, as many of the loci are neither linked directly to nor highly transcribed during the nutrient stress response (McDaniel et al., 2017). Possibly, the degenerate TATA sequence motif, which is similar to A/T-rich promoter sequences, contributes to initiation of transcription at these loci. In addition to the DNA sequence, our work suggests that the chromatin architecture determines whether cryptic transcription can initiate at these sites. We observed that H3K36me1/2 and H3K36me3 are important in maintaining proper H3K27ac and H3K56ac levels at two sites of cryptic transcription. These results connect initiation of cryptic transcription to known Set2 functions – Rpd3S recruitment and activation and repelling Asf1 – and provide mechanistic understanding of how Set2-mediated H3K36me prevents cryptic transcription.

There is robust evidence that different methylation states at other histone lysine residues, such as H3K4, H3K9, and H3K27, have unique functions, but this information was lacking for H3K36 (Hyun et al., 2017). We have investigated the functions of different H3K36 methylation states *in vivo*, particularly H3K36me3. Many studies have examined the consequences of the complete absence of H3K36me or the absence of one or more methylation

states, however there has been no comprehensive analysis of the function of each H3K36 methylation state (Gopalakrishnan et al., 2019; Leung et al., 2019; McDaniel and Strahl, 2017). The distinct activities of H3K36 methylation in certain contexts and interchangeable functions in others presented in this study advances our understanding of how H3K36me contributes to transcriptional control, particularly in repressing cryptic transcription during the nutrient stress response. Our findings establish a foundation for elucidating mechanisms for how the H3K36me states contribute to a variety of cellular processes conserved across eukaryotes that are important for gene expression, development, and disease.

Materials and Methods

Yeast Strains and Plasmids

Yeast strains were created using Delitto Perfetto (Stuckey and Storici, 2013) or the PCR Toolbox (Janke et al., 2004). Yeast strains used in this study are listed in Table 2.2 and were created using the primers in Table 2.3. To construct the FLAG-tagged wild-type Set2 plasmid for the baculovirus system, the *SET2* ORF was generated by standard PCR reaction from genomic DNA and subcloned into XhoI/NotI digested pBL532 (a derivative of pBacPAK8 that contains an N-terminal HIS Tag and FLAG tag separated by a TEV protease digestion site). All other plasmids were created using site-directed mutagenesis (Agilent) with the primers in Table 2.3. Plasmids used in this study are listed in Table 2.4.

Alignment and Molecular Modeling

The SETD2 crystal structure (Yang et al., 2016) was isolated from the Protein Data Bank (PDB ID 5JLB) and served as a template for I-TASSER modeling of Set2 (Zhang, 2008). Set2 was visualized in PyMOL, and mutations were modeled with MODELLER (Webb and Sali, 2016).

Western Blotting

All strains were grown to saturation in YPD before being diluted to an OD_{600} of 0.2 and grown to OD_{600} ~1 in YPD at 30°C. For nutrient deprivation experiments, cells were isolated, washed with water, and resuspended in SD media (Cheung et al., 2008). Ten ODs of cells were collected at each time point and Western blotting was performed after extraction of proteins by TCA lysis (Keogh et al., 2006a, 2006b). Lysates were separated by SDS-PAGE, transferred to PVDF membrane, and probed overnight at 4°C with Set2 (in-house), G6PDH (Sigma A9521-1VL), H3K36me1 (abcam 9048), H2K36me2 (in-house), H3K36me3 (abcam 9050), H3K27ac (Active Motif, 39133), H3K56ac (Active Motif, 39281), H3 C-terminal (EpiCypher 13-0001), or RNAPII CTD phospho Ser2 (Active Motif 61083) primary antibodies. Membranes were washed in TBS (Tris-buffered saline)-Tween (50 mM Tris-HCl, 150 mM NaCl, and 0.5% Tween 20). Membranes were incubated with HRP-conjugated anti-rabbit (GE Healthcare NA934V; 1:10,000) or HRP-conjugated anti-rat (Jackson ImmunoResearch 312-036-045; 1:10,000) antibody and probed with ECL Prime (GE Healthcare). Immunoblots were quantified with ImageJ software for two or more independent biological replicates.

Recombinant Protein Purification from Baculovirus Expression System

Ten mL of wild-type Set2, *set2-Y149F*, or *set2-F234Y* P2 virus was inoculated into 100 mL freshly conditioned Sf21 cells (1x106 cells/mL), supplemented with 10% FBS, 1% P/S, and cultured at 27°C for 48 hours. Cells were collected by centrifugation at 3,000 rpm for 5 minutes in a table-top centrifuge and washed with 10 mL 1x PBS buffer. Cell pellets were lysed in 10 mL BV-lysis buffer (50 mM HEPES pH 7.9, 500 mM NaCl, 10% glycerol, 2mM MgCl2 and 0.2% Triton) on ice for 30 minutes. The lysates were clarified by ultracentrifugation (Beckman 50.2 Ti 40,000 rpm) for 30 minutes. Flag beads (200 µl, Sigma) were added to the supernatant and incubated at 4°C for at least 2 hours. After incubation, beads were washed three times with 10 mL BV-lysis buffer and eluted with 600 µl of FLAG-elution buffer (50 mM HEPES pH 7.9, 100

mM NaCl, 2 mM MgCl2, 0.02% NP40; 10% glycerol) containing 500 µg/mL of 3xFlag peptide. Elutions were concentrated with a 10-kD cut-off concentrator (Amicon).

Histone Methyltransferase Assay

Standard HMT reactions were performed in a 40 µl system containing 1x HMT buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM MgCl2, 2 mM DTT, 5% glycerol), 4 µg or 3-fold serial dilution of recombinant Set2 protein, 4 µg of Xenopus recombinant nucleosomes supplemented with 10 µM unlabeled S-adenosylmethionine (Sigma) (Li et al., 2003). Reactions were incubated at 30oC for 1 hour. Immunoblots were performed with antibodies against H3K36me3 (abcam 9050), H3K36me2 (abcam 9049), and H3K36me1 (abcam 9048) to detect specific histone methylation states.

Co-immunoprecipitation

All strains were grown to saturation in SC-Leu before being diluted to an OD_{600} of 0.4 and grown to OD_{600} ~2.0 in 100 mL of SC-Leu at 30°C. The cells were pelleted and washed with 30 mL of distilled H_2O. The pellets were resuspended in 500 µl of Extraction Buffer (Kizer et al., 2005) and split equally into two tubes. Glass beads (BioSpec Products) were added to bring the total volume to 1 mL and samples were vortexed for 10 min at 4°C. Lysates were collected into fresh tubes via centrifugation, and the lysates were cleared at max speed for 15 min at 4°C. Protein concentration was measured via Bradford assay (Bio-Rad). A 25 µl aliquot was used for input and 1 mg/mL of protein was incubated overnight in 600 µl of Extraction Buffer at 4°C with 5 µl of HA antibody (Bethyl). Antibody was conjugated to 40 µl Protein A Agarose beads (Roche) for 2 hours at 4°C before being washed with Extraction Buffer and eluted with 50 µl of 1x SDS buffer. Samples were heated at 95°C for 5 minutes and loaded onto an 8% SDS-PAGE gel. Data are representative of two independent biological replicates.

Chromatin Immunoprecipiration and Real-Time PCR

Yeast strains were grown to saturation in YPD before being diluted to an OD_{600} of 0.2 and grown to an OD_{600} of ~1 at 30°C. For nutrient deprivation experiments, cells were isolated, washed with water, and resuspended in SD media (Cheung et al., 2008). Fifty ODs of cells were collected at each time point and ChIP was performed as described with modifications (Ahn et al., 2009). For precipitation of proteins, 2 ul of H3K36me1 (abcam 9048), H2K36me2 (Active Motif 39255), H3K36me3 (abcam 9050), H3K27ac (Active Motif, 39133), H3K56ac (Active Motif, 39281), or H3 C-terminal (EpiCyphe, 13-0001) was used and IgG (Cell Signaling 2729) was the negative control. DNA was eluted in 30 µl of elution buffer (Zymo) and diluted 1:20. Two µl of the diluted DNA was subjected to qPCR using SYBR Green (Bio-Rad) and primers described in Table 2.3. Data are represented as the mean percent input values ± standard error of the mean (SEM) from two or more biological replicates with technical duplicates.

Spotting Assays

Yeast strains were grown at 30°C to saturation and diluted to an OD_{600} of 0.5 prior to spotting 5-fold serial dilutions on plates at 30°C or 34°C, as indicated, for 2-4 days. For the Bur1 growth assay, *BUR1* deletion shuffle strains were grown on media lacking uracil to maintain the wild-type *BUR1* plasmid before plating on media containing the 5-fluroorotic acid (5-FOA). All experiments were performed at least three times.

RNA-seq Library Preparation and Sequencing

Yeast strains were grown to saturation in YPD before being diluted to an OD_{600} of 0.2 and grown to an OD_{600} of ~1 at 30°C. Cells were isolated, washed with water, and resuspended in SD media (Cheung et al., 2008). Ten ODs of cells were collected at each time point and RNA was isolated by acid phenol (Ambion/Thermo Fisher Scientific) extraction. Five µg RNA was treated with DNase (Promega) and purified (RNeasy column, QIAGEN). Two and a half µg of the purified RNA was processed with yeast-specific rRNA depletion beads (Illumina). Strand-

specific bar-coded sequencing libraries were then prepared (TruSeq Stranded Total RNA Library Preparation Kit, Illumina). Libraries were pooled and sequenced across two lanes (Hi-Seq 4000, Illumina). Data are representative of three independent biological replicates.

Sequencing Alignment and Analysis

Reads were trimmed and filtered of adapter sequences using cutadapt, and required to have at least 90% of bases with quality scores exceeding 20. Reads were aligned to the reference yeast genome (sacCer3) using STAR. Sense and anti-sense reads were analyzed separately (Dobin et al., 2013). Coverage across genomic features was computed using BEDtools (Quinlan and Hall, 2010). Significantly enriched motifs were discovered using HOMER, with CIS ± 100bp as foreground and the 100 bp flanking sequence on either side as background (Heinz et al., 2010). As a control, we randomly selected 439 low-confidence CIS and repeated the motif analysis. Signal was normalized to total sequencing depth and biological replicates were averaged where appropriate.

Tables

Table 2.1: Summary of set2 Mutant Phenotypes

	Caffeine Sensitivity	Rapamycin Sensitivity	Phleomycin Sensitivity	6AU Resistance	*bur1*Δ Bypass	*spt16-11* Bypass	Cryptic Initiation
Wild-type	-	-	-	-	-	-	-
***set2*Δ**	+	+	+	+	+	+	+
set2-Y149F	-	-	-	-	-	-	-
set2-F234Y	-	-	-	+/-	-	-	-
set2-H199L	+	+	+	+	+	+	+

A minus (-) indicates wild-type phenotype and a plus (+) indicates *set2*Δ phenotype. A plus/minus (+/-) indicates an intermediary phenotype.

Table 2.2: Yeast Strains and Genotypes

Name	Genotype	Reference
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Open Biosystems
BY4742	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	Open Biosystems
YSM244	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 set2Δ::NAT	McDaniel et al., 2017
YJD21	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 *set2-Y149F*	This study
YJD22	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 *set2-F234Y*	This study
YJD23	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 *set2-H199L*	This study
YSB788	MATalpha ura3-52 leu2Δ1 trp1Δ63 his3Δ200 bur1::*HIS3* lys2Δ202 (pRS316-Bur1)	M.C. Keogh et al., 2003
YSB1003	MATalpha ura3-52 leu2Δ1 trp1Δ63 his3Δ200 bur1::*HIS3* lys2Δ202 set2::KanMX (pRS316-Bur1)	Michael Christopher Keogh et al., 2005
DY7230	MATa ade2 can1 his3 lue2 lys2 met15 ura3 *spt16-11*	Voth et al., 2014
YSM174	MATa ade2 can1 his3 leu2 lys2 met15 ura3 *spt16-11* set2::HYGRO	This study
KLY78	MATalpha ura3-52 leu2Δ1 trp1Δ63 his3Δ200 KanMX6::GAL1pr-flo8::*HIS3* lys2-128δ	Nourani, Robert, & Winston, 2006
YSM138	MATalpha ura3-52 leu2Δ1 trp1Δ63 his3Δ200 KanMX6::GAL1pr-flo8::*HIS3* lys2-128δ set2::NatMX	This study
YBL860	MATa his3Δ1, leu2Δ0 met15Δ0 ura3Δ0 *STE11* -1870::*HIS3*	Wang et al., 2015
YBL861	MATa his3Δ1, leu2Δ0 met15Δ0 ura3Δ0 set2Δ::HYGRO *STE11* -1870::*HIS3*	Wang et al., 2015

Table 2.3: List of Primers

Name	Sequence	Target	Source
oJD08	CGATCACCTCACCTTTGAACTCGTAAATGAACTGG	Forward primer *set2-Y149F* site directed mutagenesis	This study
oJD09	CCAGTTCATTTACGAGTTCAAAGGTGAGGTGATCG	Reverse primer *set2-Y149F* site directed mutagenesis	This study
oJD10	CATAACGATCCACATTATAATCATAAGTGATTTCCTCACCTTTTAAAATTTTTCTT	Forward primer *set2-F234Y* site directed mutagenesis	This study
oJD11	AAGAAAAATTTTAAAAGGTGAGGAAATCACTTATGATTATAATGTGGATCGTTATG	Reverse primer *set2-F234Y* site directed mutagenesis	This study
H199LFor	GGCCAGATTTTGCAATCTCTCTTGCAGCCCCAATG	Forward primer *set2-H199L* site directed mutagenesis	This study
H199LRev	CATTGGGGCTGCAAGAGAGATTGCAAAATCTGGCC	Reverse primer *set2-H199L* site directed mutagenesis	This study
oJD24	GAGCCGAACAGGACATAGAAGCCAACCAGTTCATTTACGAGAGCTCGTTTTCGACACTGG	*set2-Y149F* P.1 for Delitto Perfetto	This study
oJD25	AACCTATCTCTAAATTCCATCTCCTCGATCACCTCACCTT TCCTTACCATTAAGTTGATC	*set2-Y149F* P.2 for Delitto Perfetto	This study
oJD26	TATTTGCTCAAAGAAAAATTTTAAAAGGTGAGGAAATCACGAGCTCGTTTTCGACACTGG	*set2-F234Y* P.1 for Delitto Perfetto	This study
oJD27	CATTTCTGAGCTTGAGCACCATAACGATCCACATTATAATTCCTTACCATTAAGTTGATC	*set2-F234Y* P.2 for Delitto Perfetto	This study
oJD28	GAGCCGAACAGGACATAGAAGCCAACCAGTTCATTTACGAGTTTAAAGGTGAGGTGATCGAGGAGATGGAATTTAGAGATAGGTT	*set2-Y149F* IRO for Delitto Perfetto	This study
oJD29	TATTTGCTCAAAGAAAAATTTTAAAAGGTGAGGAAATCACTTACGATTATAATGTGGATCGTTATGGTGCTCAAGCTCAGAAATG	*set2-F234Y* IRO for Delitto Perfetto	This study
oJD38	TTGACGCCACAATAAAGGGTTCGTTGGCCAGATTTTGCAATTTGTCTTGCAGCCCCAATGCATATGTTAATAAATGGGTTGTTAA	*set2-H199L* IRO for Delitto Perfetto	This study
oJD45	GGAGAATTCATTGACGCCACAATAAAGGGTTCGTTGGCCAGATTTTGCAAGAGCTCGTTT	*set2-H199L* P.1 for Delitto	This study

	TCGACACTGG	Perfetto	
oJD46	TAGCTTATCCTTAACAACCCATTTATTAACA TATGCATTGGGGCTGCAAGTCCTTACCATT AAGTTGATC	*set2-H199L* P.2 for Delitto Perfetto	This study
oJD69	ATCTTGCACTGGGTCTTACG	*STE11* ChIP	This study
oJD70	AGGAAAAGTTGGGCGAGAATC	*STE11* ChIP	This study
oJD71	CATGGAACAGACACAAACAGC	*STE11* ChIP	This study
oJD72	TGCACAAAAGGTAAATCGTTGG	*STE11* ChIP	This study
oJD73	CGGATTGAACAGGTGAATAGATTG	*STE11* ChIP	This study
oJD74	AGCGGAACCATCGTTTAAGAT	*STE11* ChIP	This study
oJD75	CCCACAGACAACAATAAACAAGC	*STE11* ChIP	This study
oJD76	GCTGTAAAGCATCAACCATCTTT	*STE11* ChIP	This study
oJD77	GATATCAAAGGTTGCGTAAAAATTACTG	*STE11* ChIP	This study
oJD78	GCAGTAGTAGCGGTCTGTTTG	*STE11* ChIP	This study
oJD79	CTAGTGCCCTTGAATTGCTG	*STE11* ChIP	This study
oJD80	AATCGGCCAGAGCACTTTAG	*STE11* ChIP	This study
oJD81	CTCATACCCCTGGAATAAAATCAAG	*STE11* ChIP	This study
oJD82	GAAAGGCCTGTTTCTTCGTG	*STE11* ChIP	This study
PMA1	CCCTCGTTCACAGAAAGTCTGAAGAAG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA2	GGAGCATAAGCGGTACCCACC	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA3	CCCCAGCTAGTTAAAGAAAATCATTGAAAA G	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA4	CTTAGCAGGCTTTTCTTGAGTTGGC	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA5	GTACGGTTTGAATCAAATGGCTG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA6	CACAGATAACACCGAAATCGACCC	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA7	CGATCAATCTGCTATTACTGGTG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA8	CCAAAGCAGCAGCTCTACCAACGAAAG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015

PMA9	CTTGGGTCTATGGATTGCTATTTTGG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA1 0	CTTGGTAGGTTCCATTTAACGGGCTTTGG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA1 1	TGGTGGTTTCTACTACGAAATGTCC	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA1 2	TGATTAAATGCTACTTCAACAGGATTAGG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA1 3	GCCAACAAGAATAAGCCGCTTATTTCC	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA1 4	CAAATAAAACAACCAGCTTCGGTGTGTG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA1 5	GCCTCCGCGAAATACCTTTACTGATTTTG	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
PMA1 6	AACTGAGTCATCTAGAGTAATGACGC	*PMA1* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH1	CACAACCTCAATGGAGTGATGC	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH2	GGGGAATAATTTCAGGGAACTGG	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH3	TTAGAGTTGCTATTAACGGTTTCGG	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH4	CATCGTGGGAAACTTCACCAGC	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH5	CTTGTACCACCAACTGTTTGGC	*TDH3* ChIP	Grzechnik, Gdula, &

			Proudfoot, 2015
THD6	ATGGGATGATGTTACCGGAAGC	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH7	GAAGACGCTGTTGTCTCCTCTG	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH8	AAGCCTTGGCAACGTGTTCAAC	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH9	CTATTTTAATGACATTTTCGATTCATTG	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH10	CCAAAATTATTAAGAGCGCCTCC	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH11	GCCTATAAATCATGCCTATATTTGCG	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
TDH12	CCGCGGGAATCTGTGTATATTAC	*TDH3* ChIP	Grzechnik, Gdula, & Proudfoot, 2015
oJD119	TGTTATCCTCGATGTGCTGGA	*SPB4* ChIP	This study
oJD120	AGCGATTTCAACAGGTACCAA	*SPB4* ChIP	This study
oJD121	TGAGCTGAGCCCATTGATTTG	*SPB4* ChIP	This study
oJD122	GCTTGGACTGGAGTCATGGT	*SPB4* ChIP	This study
oJD123	TCCACTCGTTAATTATTGCTCCA	*SPB4* ChIP	This study
oJD124	AGTACCAACCAGGAGCTGAC	*SPB4* ChIP	This study
oJD125	TCAAACATCGGCAAGGACAA	*SPB4* ChIP	This study
oJD126	GCTGGATGACGAGATCAACG	*SPB4* ChIP	This study
oJD127	ACTCAAATCATTCCGCAACTTCA	*SPB4* ChIP	This study
oJD128	GGATCGACCAACCAATTCCC	*SPB4* ChIP	This study
oJD129	TCAGCATGTACTTGTAACAGGA	*SPB4* ChIP	This study
oJD130	TGTGCCATCTACTCACACTCA	*SPB4* ChIP	This study

Table 2.4: List of Plasmids

Name	Features	Description	Reference
pCORE	kanMX4, KIURA3, Amp	CORE plasmid for Delitto Perfetto	Stuckey & Storici, 2013
pBL532-Set2-Flag	Amp	*SET2-Flag*	This study
pBL532-*set2-Y149F* -Flag	Amp	*set2-Y149F* -Flag	This study
pBL532-*set2-F234Y* -Flag	Amp	*set2-F234Y* -Flag	This study
pRS415	CEN6_ARS4, LEU2, Amp	Empty Vector	Mumberg, Mailer, & Funk, 1995
pRS415-Set2-HA	CEN6_ARS4, LEU2, Amp	*SET2*-HA	Du et al., 2008
pJD01	CEN6_ARS4, LEU2, Amp	*set2-Y149F* -HA	This study
pJD02	CEN6_ARS4, LEU2, Amp	*set2-F234Y* -HA	This study
pJD03	CEN6_ARS4, LEU2, Amp	*set2-H199L* -HA	This study
p416	CEN6_ARS4, URA3, Amp	Empty Vector	Mumberg et al., 1995
p416-ADH1-Set2	CEN6_ARS4, URA3, Amp	*SET2*-Flag	Brian D Strahl et al., 2002
pJD04	CEN6_ARS4, URA3, Amp	*set2-Y149F* -3xFlag	This study
pJD05	CEN6_ARS4, URA3, Amp	*set2-F234Y* -3xFlag	This study
pJD06	CEN6_ARS4, URA3, Amp	*set2-H199L* -3xFlag	This study
pRS316 Bur1	CEN6_ARS4, URA3, Amp	*BUR1*	M.C. Keogh et al., 2003

Figures

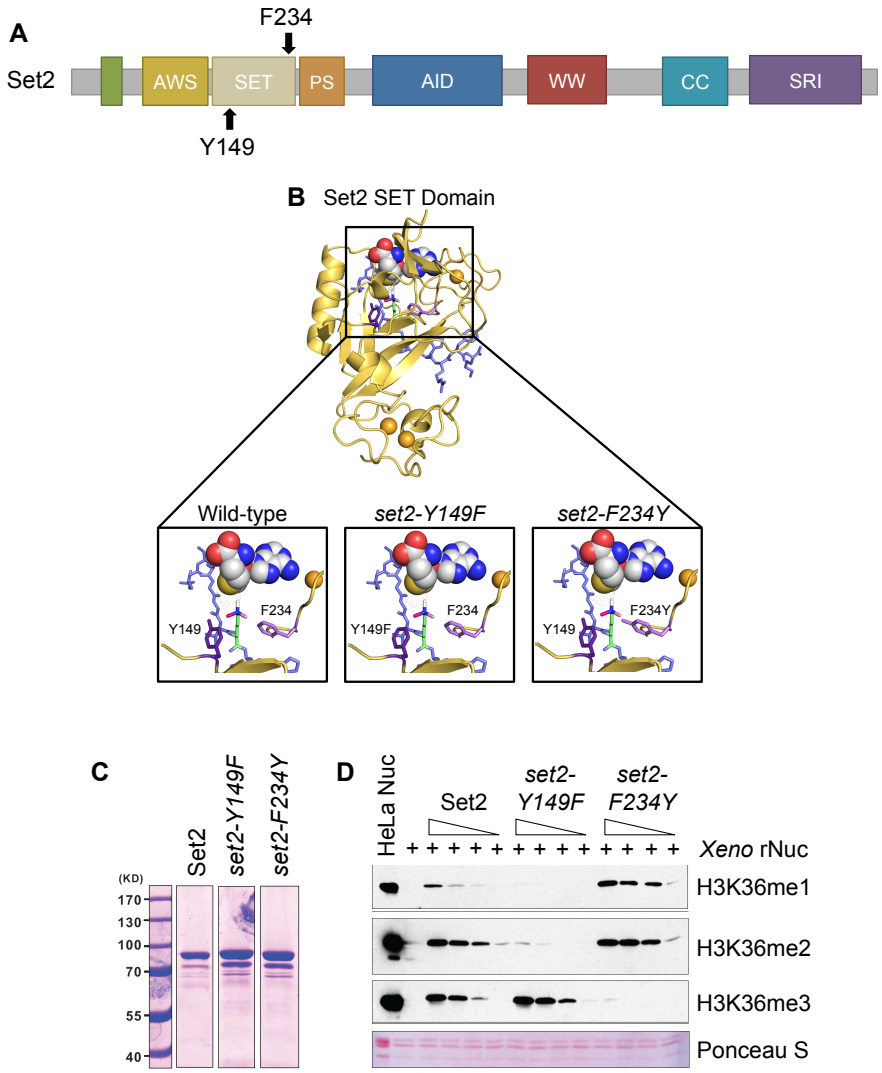


Figure 2.1: Phe/Tyr Switch in Set2 Separates H3K36 Methylation States *in vitro*

A. Domain map of Set2 with the Y149 and F234 residues of the Phe/Tyr switch highlighted. The catalytic region of Set2 is composed of the associated with SET (AWS), catalytic SET domain, and post-SET (PS). The autoinhibitory domain (AID) regulates catalytic activity. The C-terminus of Set2 has protein-protein interaction domains, like the coiled-coiled (CC), WW, and Set2-Rpb1 (SRI) protein-protein interaction domain. B. Model of Set2 SET domain (yellow) bound to H3 peptide (slate) and SAM (yellow, white, blue, and red spheres) with locations of the F234 (light purple) and Y149 (dark purple) that form the Phe/Tyr switch highlighted. H3K36 is green with methylation shown in white, light pink, and dark pink. Orange spheres are zinc ions necessary for catalysis. C. Coomassie staining of recombinant Set2 and Phe/Tyr switch mutant proteins. D. Western blots of *in vitro* histone methyltransferase (HMT) assays using the indicated antibodies. In vitro HMT assays were performed with an equal amount of recombinant Set2 and Phe/Tyr switch mutant proteins from insect cells, recombinant Xenopus nucleosomes, and co-factor S-adenosylmethionine (SAM). HeLa LON (long oligonucleosomes) were used as a positive control for Western blotting.

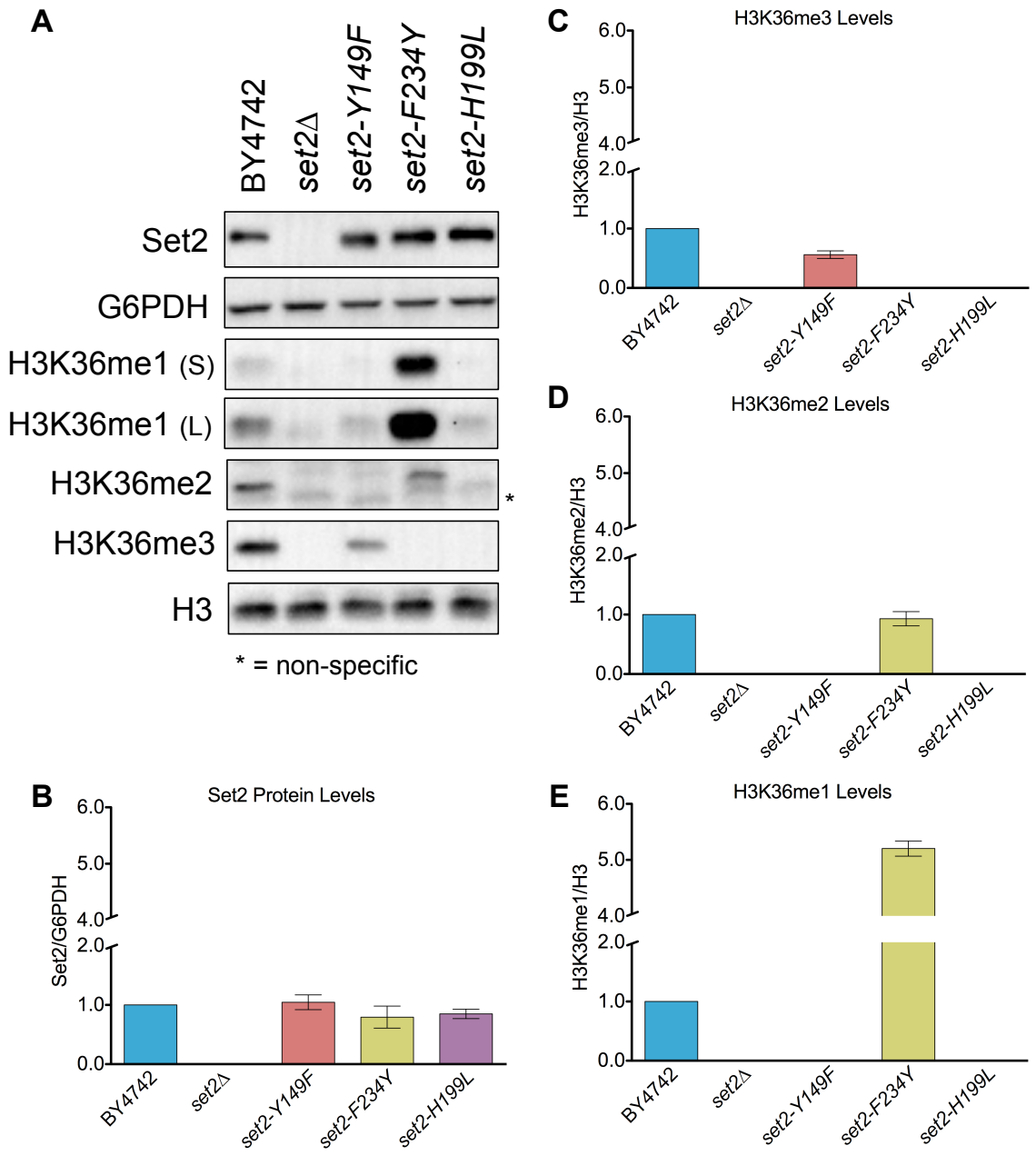


Figure 2.2: Phe/Tyr Switch Mutations in Set2 are Tools to Separate H3K36 Methylation States *in vivo*

A. Western blots of yeast strains probed with Set2 and different H3K36 methylation antibodies. S = short exposure and L = long exposure. G6PDH and H3 served as loading controls. B. Quantification of Set2 normalized to G6PDH in the indicated strains. Data for mutants were normalized relative to BY4742. C-E. Quantification of H3K36me3 (C), H3K36me2 (D), and H3K36me1 (E) normalized to H3 in the indicated strains. H3K36me1 was quantified using the short exposure Western blot. Measurement for all mutants were normalized BY4742. Each bar graph is representative of mean ± SEM of two or more independent biological replicates with a representative replicate shown in (A).

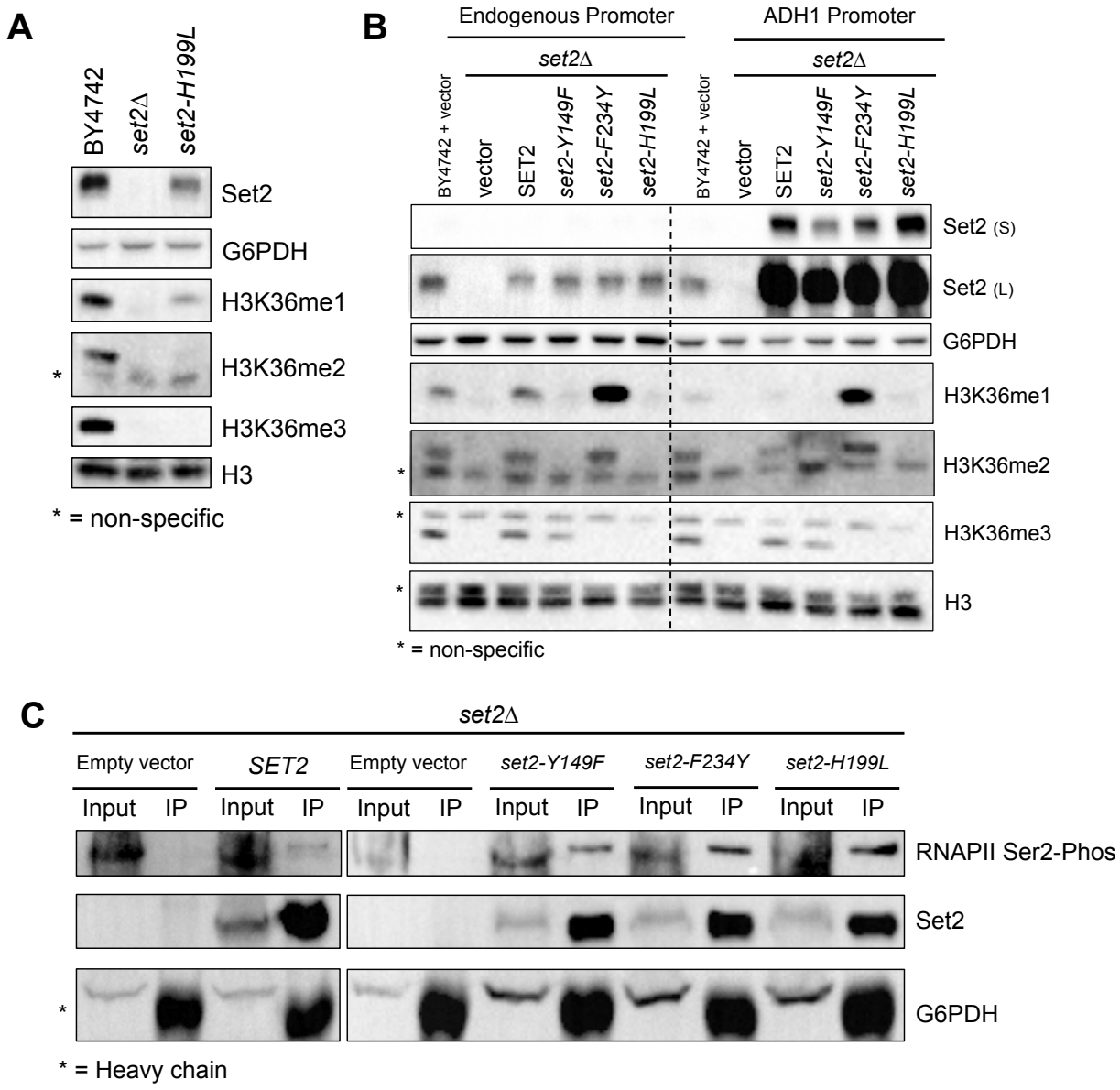


Figure 2.3: *set2* Mutatns Differentially Methylate H3K36 and Interact with RNAPII

A. Western blots of indicated strains probed with Set2 and different H3K36 methylation antibodies. H3 and G6PDH served as loading controls B. Western blots of indicated strains transformed with plasmids that expressed *SET2* from its native promoter or overexpressed *SET2* from the *ADH1* promoter and probed with Set2 and different H3K36 methylation antibodies. Loading controls were H3 and G6PDH. C. Co-immunoprecipitation showing interaction of *set2* mutants and serine 2 phosphorylated form of RNAPII. Set2 was immunoprecipitated by anti-HA antibody. G6PDH served as a loading control.

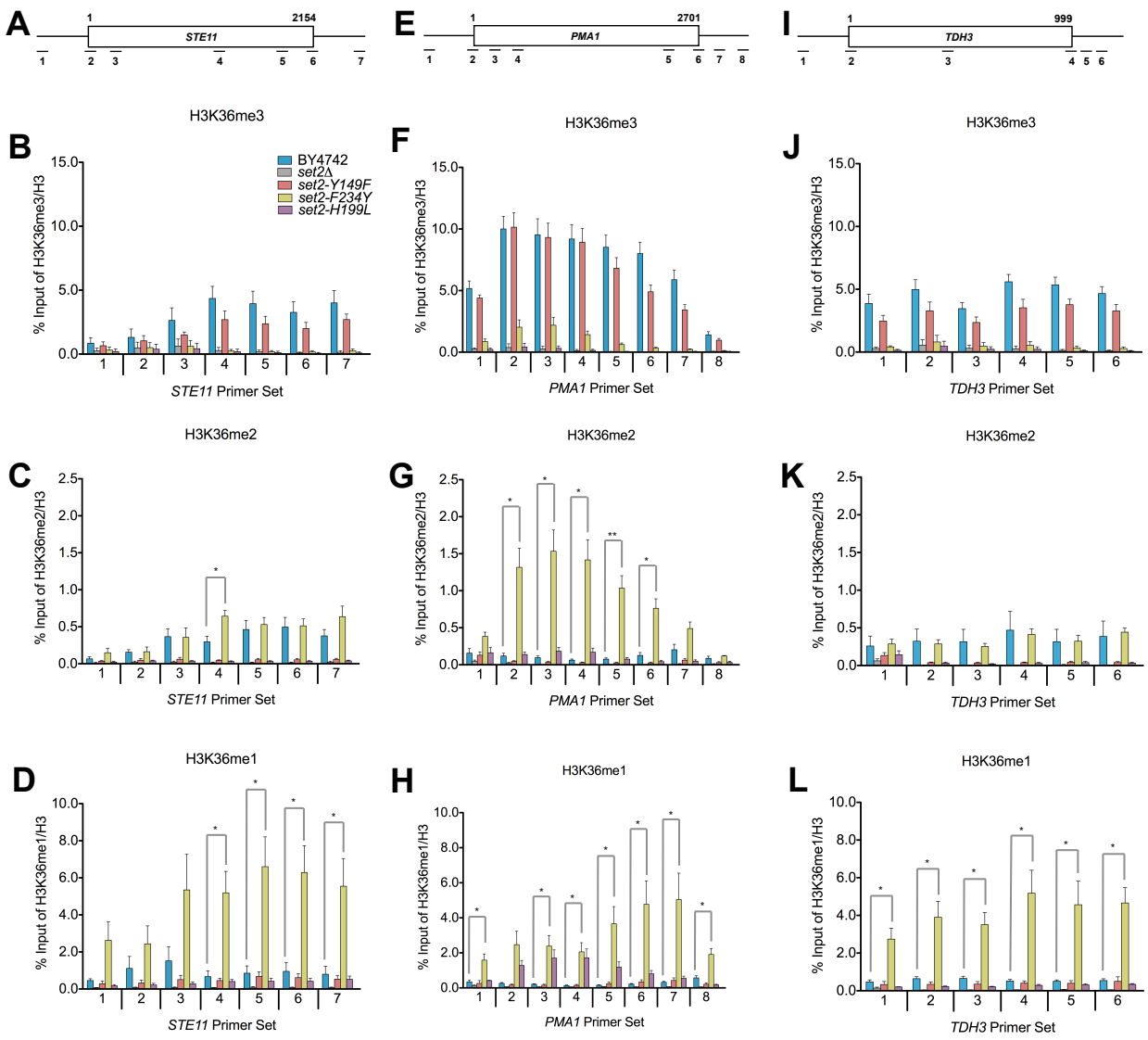


Figure 2.4: Distinct Methylated Forms of H3K36 are Deposited within or near Transcribed Regions of Genes

A. Schematic of *STE11* with amplicons indicated below. B-D. ChIP analysis of H3K36me3, H3K36me2, and H3K36me1 across *STE11* in the indicated strains. The legend in (B) is representative for all the data presented in the Figure. E. Schematic of *PMA1* with amplicons indicated below. F-H. ChIP analysis of H3K36me3, H3K36me2, and H3K36me1 across *PMA1* in the indicated strains. I. Schematic of *TDH3* with amplicons indicated below. J-L. ChIP analysis of H3K36me3, H3K36me2, and H3K36me1 across *TDH3* in the indicated strains. Data represented as mean ± SEM of three independent biological replicates. Student's t-test was used to obtain p-values. Asterisks indicate significance (* $p<0.05$; ** $p<0.01$); non-significant comparisons not shown. All qPCR primers are listed in Table 2.3.

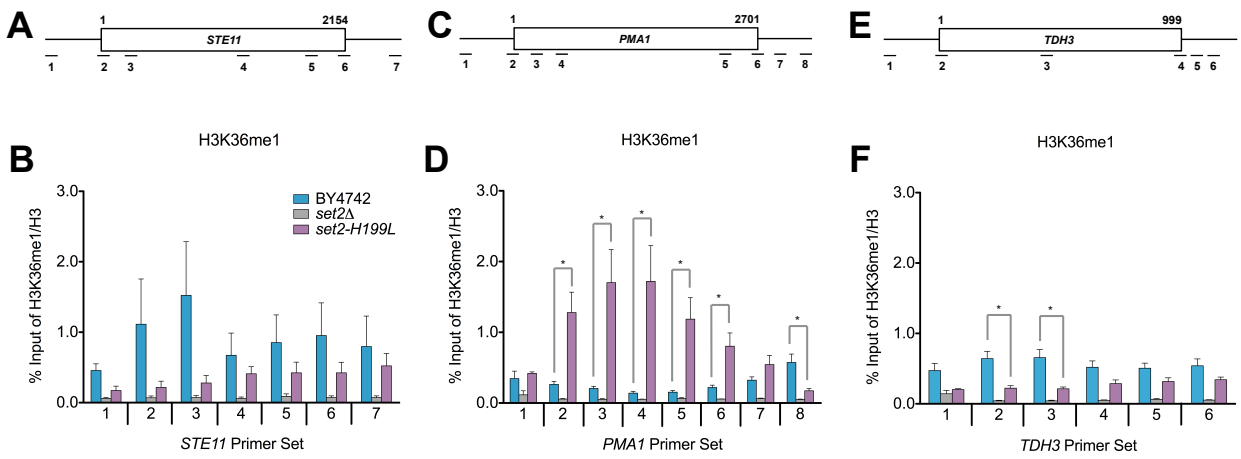


Figure 2.5: H3K36me1 is Deposited within or near Transcribed Regions of Genes

A. Schematic of *STE11* with amplicons indicated below. B. ChIP analysis of H3K36me1 across *STE11* in the indicated strains. C. Schematic of *PMA1* with amplicons indicated below. D. ChIP analysis of H3K36me1 across *PMA1* in the indicated strains. E. Schematic of *TDH3* with amplicons indicated below. F. ChIP analysis of H3K36me1 across *TDH3* in the indicated strains. Data represented as mean ± SEM of three independent biological replicates. Student's t-test was used to obtain p-values. Asterisks indicate significance (* $p<0.05$); non-significant comparisons are not shown. All qPCR primers are listed in Table 2.3.

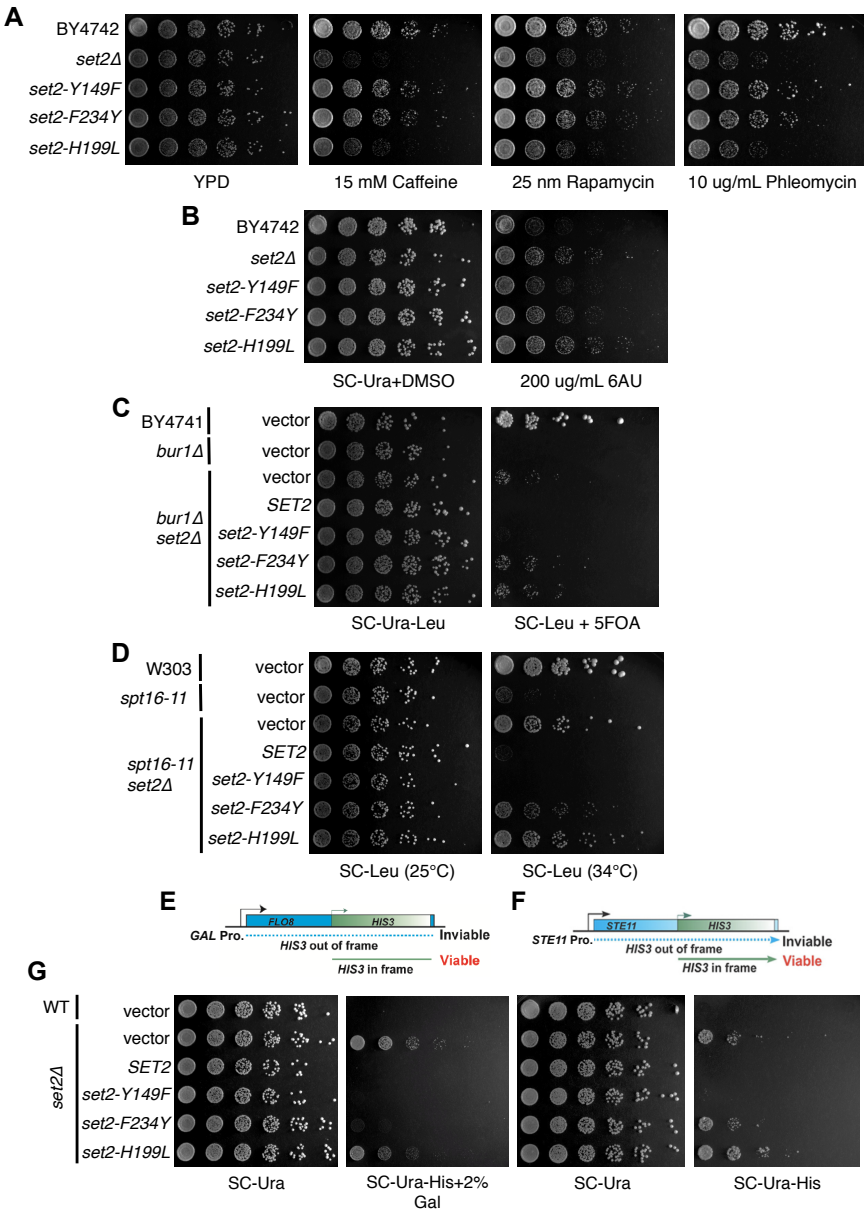


Figure 2.6: H3K36me1/2 and H3K36me3 have Unique Phenotypes in Some Cellular Contexts

A. Five-fold serial dilutions of BY4742, *set2*Δ, and set2 mutant strains plated on YPD or YPD containing caffeine (15 mM), rapamycin (25 nM), or phleomycin (10 ug/mL). B. Five-fold serial dilutions of BY4742, *set2*Δ, and *set2* mutant strains plated on SC-Ura containing DMSO or 200 µg/mL 6AU. C. Five-fold serial dilutions of BY4741, *bur1*Δ , and set2 mutant strains plated on SC-Ura-Leu or SC-Ura-Leu containing 5-FOA. D. Five-fold serial dilutions of W303, *spt16-11,* and set2 mutant strains plated on SC-Leu and incubated at 25°C or 34°C. E. Schematic of *FLO8 -HIS3* fusion gene reporter to detect cryptic transcription. F. Schematic of *STE11 -HIS3* fusion gene reporter to detect cryptic transcription. G. Five-fold serial dilutions of indicated wild-type, *set2*Δ, and *set2* mutant strains plated on SC-Ura, SC-Ura-His with 2% galactose, or SC-Ura-His. All spotting assays were repeated three times and the images shown are representative of the data.

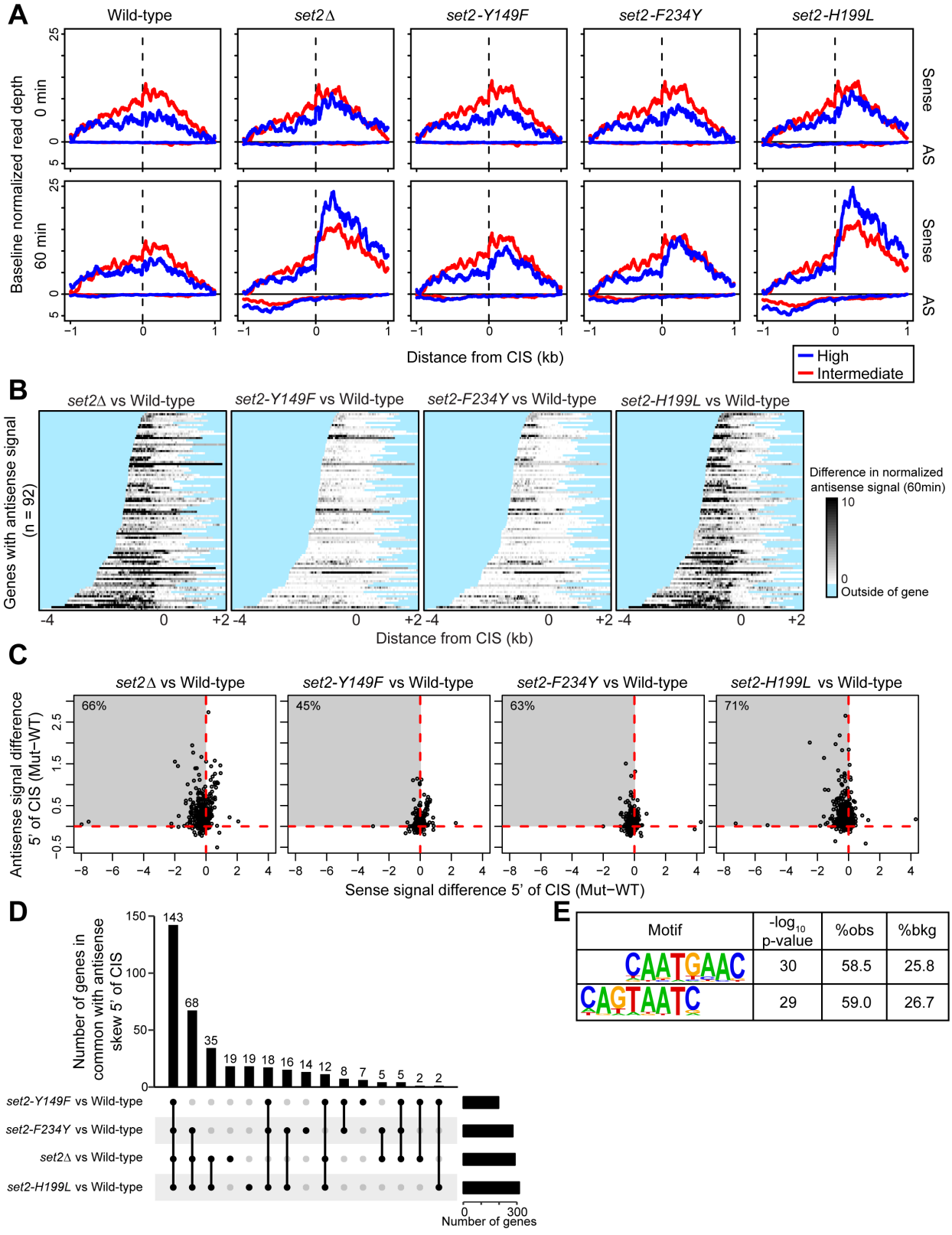
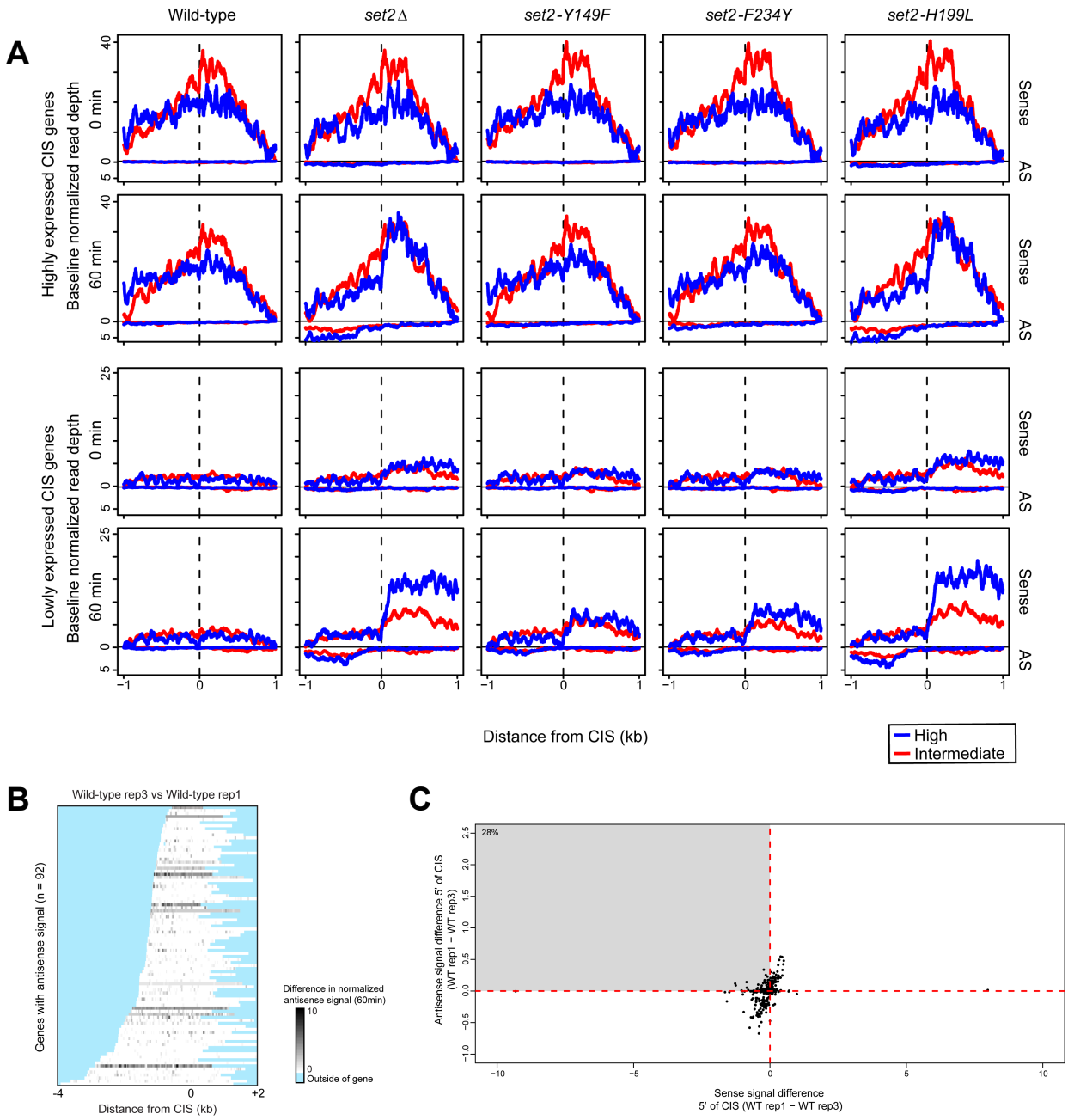

A
Wild-type
set2Δ
set2-Y149F
set2-F234Y
set2-H199L
Baseline normalized read depth
0 min
60 min
Sense
AS
Distance from CIS (kb)
High
Intermediate
B
set2Δ vs Wild-type
set2-Y149F vs Wild-type
set2-F234Y vs Wild-type
set2-H199L vs Wild-type
Genes with antisense signal (n = 92)
Distance from CIS (kb)
Difference in normalized antisense signal (60min)
10
0
Outside of gene
C
set2Δ vs Wild-type
set2-Y149F vs Wild-type
set2-F234Y vs Wild-type
set2-H199L vs Wild-type
66%
45%
63%
71%
Antisense signal difference 5' of CIS (Mut−WT)
Sense signal difference 5' of CIS (Mut−WT)
D
Number of genes in common with antisense skew 5' of CIS
143
68
35
19
19
18
16
14
12
8
7
5
5
2
2
set2-Y149F vs Wild-type
set2-F234Y vs Wild-type
set2Δ vs Wild-type
set2-H199L vs Wild-type
Number of genes
E
Motif
-log10 p-value
%obs
%bkg
30
58.5
25.8
29
59.0
26.7

Figure 2.7: Function of H3K36 Methylation States in Cryptic Transcription Regulation

A. Sense and antisense normalized transcriptional signal (reads per million mapped) across 439 high- (blue) and intermediate- (red) confidence cryptic initiation sites (CIS) defined previously. Signal was averaged across three independent biological replicates and plotted for 0 min (top) and 60 min (bottom) following nutrient deprivation for each genetic model. The minimum value for each line was adjusted to zero (y axis) to adjust for subtle differences in baseline expression and to focus on the position and range in magnitude of signal (see McDaniel et al., 2017). B. Heatmap of antisense transcription, plotted as the difference in antisense signal between set2 mutant and wild-type (mutant–WT) at 60 min following nutrient deprivation. Normalized signal is plotted for 92 genes shown previously to have antisense transcription between the CIS and transcription start site. Darker gray indicates more antisense signal in mutant than in wild-type. Regions outside of the gene body are masked blue. C. Scatterplot of sense and antisense signal differences (mutant–wild-type) in the mean per-base coverage over the gene regions between the CIS and transcription start site, for the 439 high- and intermediate-confidence CIS. Each point represents the CIS of a given gene; points that extend into the upper-left quadrant indicate decrease in sense transcription (relative to wild-type) and a concomitant increase in antisense transcription. The percentage of all 439 genes falling in this quadrant is supplied in the top left of each panel. D. UpSet plot showing overlaps for genes in the upper-left quadrants of C. One hundred and forty-three genes exhibited this antisense skew relative to wild-type in all four Set2 models. E. Significantly enriched sequence motifs discovered in the 100 bp surrounding the 439 CIS, requiring at least 2-fold enrichment relative to local background sequence.



D

Motif	-log$_{10}$ p-value	%obs	%bkg
TTCAGTTC	32	62.0	27.3

Figure 2.8: Activity of H3K36 Methylation States in Cryptic Transcription Regulation

A. Sense and antisense normalized transcriptional signal (reads per million mapped) across 439 high- (blue) and intermediate- (red) confidence cryptic initiation sites (CIS) defined previously, separated by high (top) or low (bottom) gene expression. Signal was averaged across three independent biological replicates and plotted for 0min (top) and 60min (bottom) following nutrient deprivation for each genetic model. High or low expression was defined as the top or bottom 25% of genes based on their expression in wild-type. The minimum value for each line was adjusted to zero (y axis) to adjust for subtle differences in baseline expression and focus on the position and range in magnitude of signal, as in Mcdaniel et al., 2017. B. Heatmap of anti-sense transcription, plotted as the difference in antisense signal between two wild-type replicates (WT rep3–WT rep1) at 60 min following nutrient deprivation. Normalized signal is plotted for the 92 genes that were previously shown to have antisense transcription between the CIS and transcription start site. Darker gray indicates more anti-sense signal in WT rep3 compared to WT rep1. Regions outside of the gene body are masked blue. C. Scatterplot of sense and antisense signal differences (WT replicate 1– WT replicate 3) in the gene regions between the CIS and transcription start site, for 439 high- and intermediate-confidence CIS. Each point represents the CIS of a given gene; points that extend into the upper-left quadrant indicate decrease in sense transcription (relative to wild-type) with a concomitant increase in antisense transcription. The percentage of all 439 genes falling in this quadrant is supplied in the top left. D. Significantly enriched sequence motifs discovered in the 100 bp surrounding 439 randomly-selected low-confidence CIS, requiring at least 2-fold enrichment relative to local background sequence.

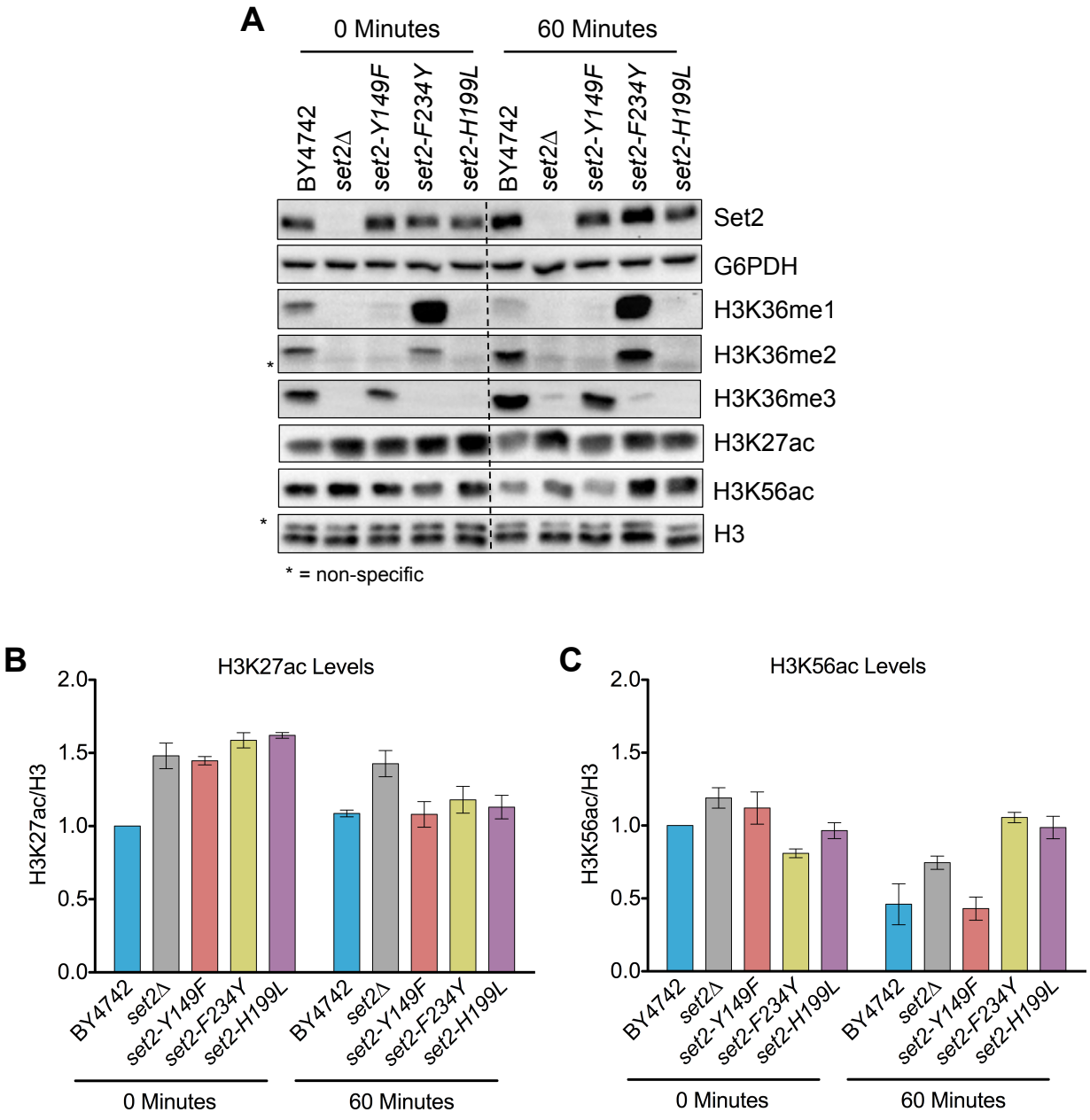


Figure 2.9: Alteration of Global H3K27ac and H3K56ac in *set2* Methylation Mutants

A. Western blots of indicated strains probed with Set2, H3K36me1, H3K36me2, H3K36me3, H3K27ac, and H3K56ac antibodies. G6PDH and H3 served as loading controls. B-C. Quantification of H3K27ac (B) and H3K56ac (C) normalized to H3. All measurements are normalized relative to BY4742 at 0 minutes. Each bar graph is representative of mean ± SEM of two or more independent biological replicates with a representative replicate shown in (A).

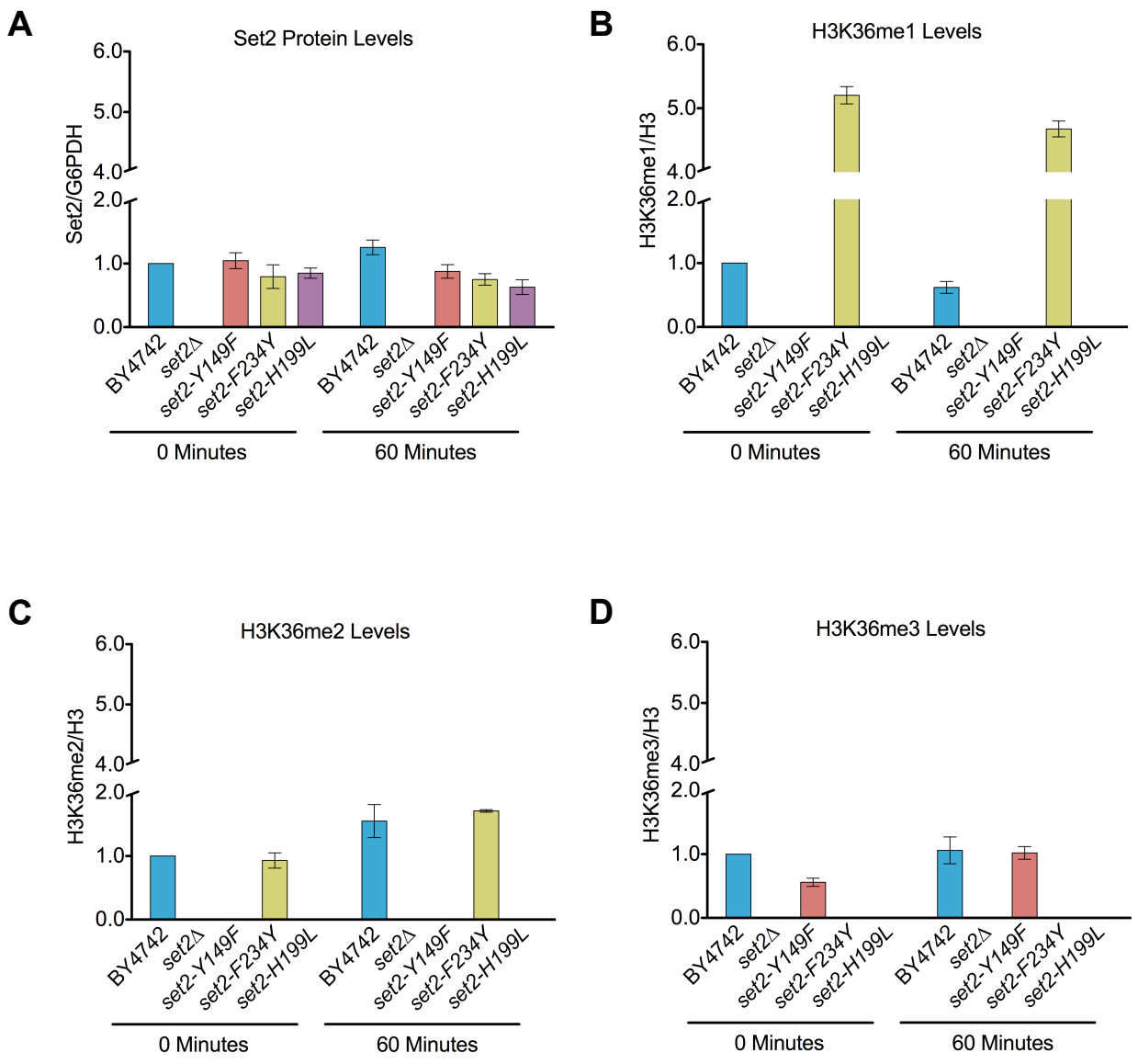


Figure 2.10: Set2 and H3K36 Methylation Levels are Similar at 0 and 60 Minutes after Nutrient Deprivation

A. Quantification of Set2 normalized to G6PDH in indicated strains. All strains were quantified relative to BY4742 at 0 minutes. B-D. Quantification of H3K36me1 (B), H3K36me2 (C), and H3K36me3 (D) normalized to H3 in the indicated strains. H3K36me1 was quantified using the short exposure Western blot. All strains were quantifed relative to BY4742 at 0 minutes. Each bar graph is representative of two or more independent biological replicates with a representative replicate shown in Figure 2.9A.

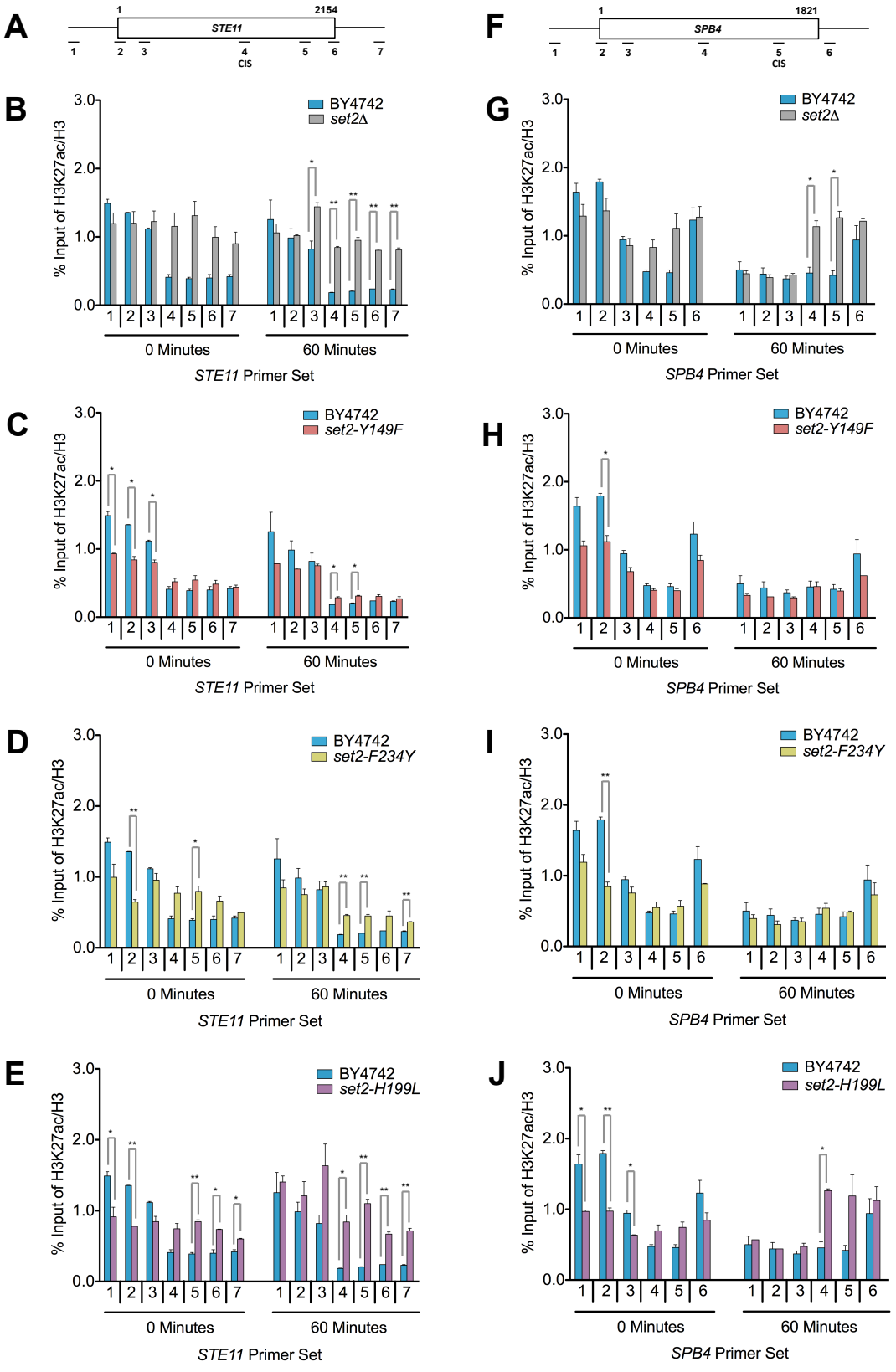

A
1
2154
STE11
1
2
3
4
5
6
7
CIS
F
1
1821
SPB4
1
2
3
4
5
6
CIS
B
BY4742
set2Δ
% Input of H3K27ac/H3
3.0
2.0
1.0
0.0
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
G
BY4742
set2Δ
% Input of H3K27ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set
C
BY4742
set2-Y149F
% Input of H3K27ac/H3
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
H
BY4742
set2-Y149F
% Input of H3K27ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set
D
BY4742
set2-F234Y
% Input of H3K27ac/H3
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
I
BY4742
set2-F234Y
% Input of H3K27ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set
E
BY4742
set2-H199L
% Input of H3K27ac/H3
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
J
BY4742
set2-H199L
% Input of H3K27ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set

Figure 2.11: H3K36me1/2 and H3K36me3 are Important for Proper H3K27ac Localization

A. Schematic of *STE11* with amplicons indicated below. Predicted cryptic initiation site (CIS) located within primer set 4. B-E. ChIP analysis of H3K27ac across *STE11* in the indicated strains and time points. F. Schematic of *SPB4* with amplicons indicated below. Predicted cryptic initiation site (CIS) located within primer set 3. G-J. ChIP analysis of H3K27ac across *SPB4* in the indicated strains and time points. Data represented as mean ± SEM of two independent biological replicates. Student's t-test was used to obtain p-values. Asterisks indicate significance (* $p<0.05$; ** $p<0.01$); non-significant comparisons are not shown. All qPCR primers are listed in Table 2.3.

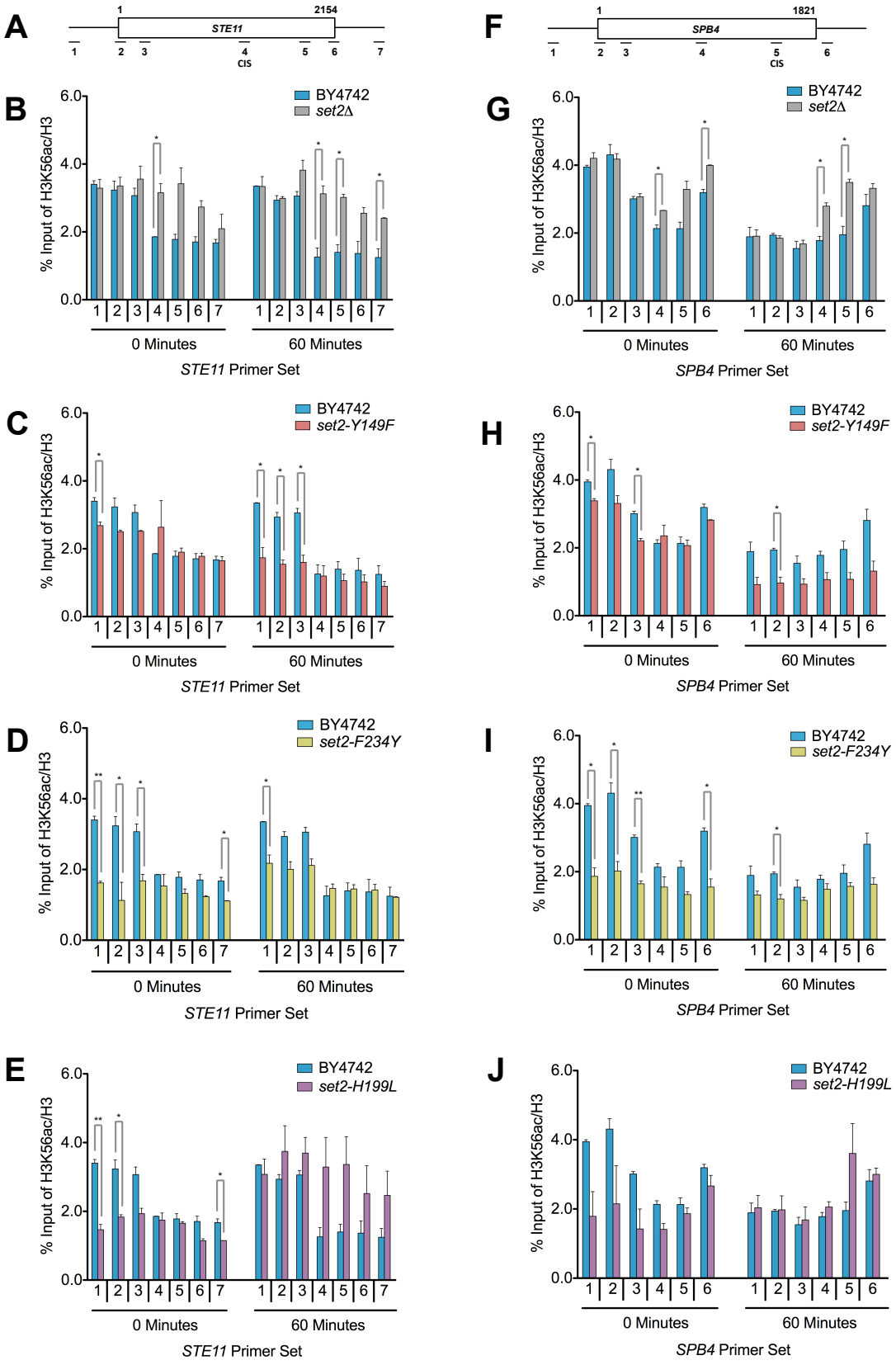

A
1
2154
STE11
1
2
3
4
CIS
5
6
7
F
1
1821
SPB4
1
2
3
4
5
CIS
6
B
BY4742
set2Δ
% Input of H3K56ac/H3
6.0
4.0
2.0
0.0
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
G
BY4742
set2Δ
% Input of H3K56ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set
C
BY4742
set2-Y149F
% Input of H3K56ac/H3
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
H
BY4742
set2-Y149F
% Input of H3K56ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set
D
BY4742
set2-F234Y
% Input of H3K56ac/H3
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
I
BY4742
set2-F234Y
% Input of H3K56ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set
E
BY4742
set2-H199L
% Input of H3K56ac/H3
1 2 3 4 5 6 7 1 2 3 4 5 6 7
0 Minutes
60 Minutes
STE11 Primer Set
J
BY4742
set2-H199L
% Input of H3K56ac/H3
1 2 3 4 5 6 1 2 3 4 5 6
0 Minutes
60 Minutes
SPB4 Primer Set

Figure 2.12: H3K26me1/2 and H3K36me3 are Important for Proper H3K56ac Localization

A. Schematic of *STE11* with amplicons indicated below. Predicted cryptic initiation site (CIS) located within primer set 4. B-E. ChIP analysis of H3K56ac across *STE11* in the indicated strains and time points. F. Schematic of *SPB4* with amplicons indicated below. Predicted cryptic initiation site (CIS) located within primer set 3. G-J. ChIP analysis of H3K56ac across *SPB4* in the indicated strains and time points. Data represented as mean ± SEM of two independent biological replicates. Student's t-test was used to obtain p-values. Asterisks indicate significance (* $p<0.05$; ** $p<0.01$); non-significant comparisons are not shown. All qPCR primers are listed in Table 2.3.

CHAPTER 3 – MODIFYING A MODIFIER: DISSECTING THE POST-TRANSLATIONAL MODIFICATION LANDSCAPE OF SET2

Summary

Set2 is a histone methyltransferase that methylates lysine 36 of histone H3 (H3K36) in the wake of RNA polymerase II. While H3K36 methylation (H3K36me) functions in transcription elongation, mRNA splicing, and DNA repair, it was not known if Set2 itself was post-translationally modified and if those modifications had biological functions. We used mass spectrometry and public databases to identify sites of methylation, acetylation, and phosphorylation on Set2. Biochemical and phenotypic analyses revealed that mutating individual sites of methylation and acetylation on Set2 does not result in obvious phenotypes, as measured by cryptic transcription initiation and transcription elongation assays. However, individually mutating three serine residues at the extreme N-terminus of Set2 that are predicted to be phosphorylated affect Set2 protein levels, but not H3K36me. Interestingly, of the six identified sites of serine phosphorylation on Set2, all of them have predicted kinases related to the cell cycle. Additionally, individually mutating each of these sites affects transcription elongation and DNA repair functions of Set2. Our results provide evidence that Set2 is post-translationally modified and sites of serine phosphorylation are functionally important.

Introduction

Protein post-translational modifications (PTMs) play a critical function in many cellular processes. While histone PTMs have been extensively studied, non-histone PTMs have recently emerged as important regulators of transcription, enzymatic activity, and protein

stability, among other functions (Biggar and Li, 2015; Narita et al., 2019; Ubersax and Ferrell, 2007). PTMs, such as methylation, acetylation, and phosphorylation, are found on thousands of non-histone proteins across many species (Duan and Walther, 2015). Analogous to how mRNA alternative splicing increases the potential products of one gene, PTMs expand the functional proteome. Perturbation of non-histone PTMs can have dire consequences for the cell, such as aberrant gene expression and improper cellular signaling, which can ultimately lead to cancer and other diseases (Biggar and Li, 2015; Narita et al., 2019; Pawson and Scott, 2005). However, many non-histone PTMs and their functions remain poorly characterized.

Lysine methylation and acetylation are PTMs that have recently emerged as important regulators of cellular function. With advances in mass-spectrometry, many sites of lysine methylation and acetylation have been identified on non-histone proteins, in particular chromatin-modifying enzymes. When these enzymes are methylated or acetylated, their protein interactions or activity can change (Biggar and Li, 2015). For example, in mice and humans, G9a automethylates itself, which serves as a binding site for HP1 proteins (Chin et al., 2007; Sampath et al., 2007). In *Saccharomyces cerevisiae* (hereafter, budding yeast), Rsc4 is acetylated by Gcn5, which inhibits normal Rsc4 binding to H3K14ac (VanDemark et al., 2007). Overall, the varied effects of lysine methylation and acetylation on chromatin-modifying enzymes in different organisms provide another layer of regulation to DNA-templated processes.

An additional and more established PTM is phosphorylation. In eukaryotes, serine, threonine, and tyrosine residues are phosphorylated and are important in signal transduction and regulating protein-protein interactions, amongst other activities (Pawson and Scott, 2005). Phosphorylated proteins are commonly at the center of protein networks and have more protein-protein interactions than non-phosphorylated proteins (Duan and Walther, 2015). Several phosphorylation-mediated protein-protein interactions involving chromatin-modifying enzymes have recently been demonstrated to be crucial for maintaining gene expression. Casein kinase

II phosphorylation of Spt6 mediates the Spt6-Spn1 interaction (Dronamraju et al., 2018). Without the Spt6-Spn1 interaction, nucleosome occupancy is disrupted and cryptic transcription occurs. Additionally, phosphorylation of the NuA4 complex is necessary for its interaction with the MRX complex during DNA damage repair (Cheng et al., 2018). Once NuA4 is recruited to sites of DNA damage via its MRX interaction, it acetylates the DNA and recruits Rpa1. The effect of phosphorylation in regulating the activity of chromatin-modifying enzymes demonstrates the importance of understanding the crosstalk between cell signaling and gene expression.

One enzyme with a critical role in gene expression is Set2, which is a histone methyltransferase. In budding yeast, it is the sole enzyme responsible for H3K36 methylation (mono-, di-, and trimethylation) (Strahl et al., 2002). During transcriptional elongation, Set2 associates with the phosphorylated C-terminal domain of RNA polymerase II and catalyzes H3K36 methylation (H3K36me) in gene bodies (Kizer et al., 2005). The downstream consequences of the histone modification deposited by Set2 are well studied and include the activation of the histone deacetylase complex (HDAC) Rpd3S, recruitment of the nucleosome remodeling complex Isw1b, and repelling the histone chaperone Asf1 (McDaniel and Strahl, 2017; Venkatesh and Workman, 2013). Taken together, these mechanisms help maintain chromatin integrity during gene transcription. However, little is known about the PTMs on Set2 itself and how those PTMs may affect the regulation and function of Set2. Other chromatin modifying enzymes, such as Rsc4 and Spt6 have biologically significant PTMs, therefore the PTMs on Set2 may also play an important function in its biology (Dronamraju et al., 2018; VanDemark et al., 2007).

In this report, we identified several residues on Set2 that are methylated, acetylated, or phosphorylated and examined their functional significance. While nine lysine residues were identified as potential sites of methylation and/or acetylation, individually mutating these residues so they could no longer be modified did not affect Set2 protein abundance or

H3K36me. Additionally, strains with the *set2* lysine point mutants phenocopied wild-type in spotting assays testing cryptic transcription suppression, transcriptional elongation, and DNA damage repair. In contrast, several sites of serine phosphorylation were identified and had functional consequences. *set2-S6A* and *set2-S8A* showed decreased Set2 protein levels compared to wild-type, while *set2-S10A* had increased Set2 protein levels. Interestingly, H3K36me was unaffected in all *set2* serine point mutants, suggesting changes in functions unrelated to H3K36 methylation. Furthermore, the individual *set2* serine point mutants could repress cryptic transcription, but showed transcription elongation defects on 6AU as well as a defect in DNA damage repair as assayed by phleomycin sensitivity. Taken together, these data provide evidence that Set2 is post-translationally modified and sites of serine phosphorylation are functionally important.

Results

Lysine Residues on Set2 are Methylated and Acetylated

The downstream consequences of the histone modification laid down by Set2 are well characterized, however, little is known about the PTMs on Set2 itself. In order to better understand the PTM landscape of Set2 and how these sites may affect Set2 regulation and function, we took an unbiased approach and used mass spectrometry to identify modified residues on Set2. Using MUD-PIT, we identified nine residues that are methylated or acetylated (Table 3.1). These residues are distributed across the enzyme, with four residues occupying the catalytic region, two in the AID, one in the coiled-coiled domain, and two in unstructured regions of the protein (Figure 3.1A). Out of the nine modified residues, four were identified as sites of dimethylation. Because trimethylation and acetylation have the same spectral count, another four sites were identified as trimethylation/acetylation and one site was identified as dimethylation and trimethylation/acetylation. Overall, these results demonstrated that Set2 is

post-translationally modified with methylation and potentially acetylation at several residues across the enzyme.

set2 Lysine Point Mutants Do Not Affect Set2 Protein Abundance or H3K36 Methylation

To test the effect of methylation and acetylation PTMs on Set2 and H3K36me *in vivo*, we generated plasmids with a lysine to alanine point mutation for each of the residues identified in Table 1 and transformed them into a strain completely lacking *SET2* (set2Δ). For controls, we included a *set2*Δ strain, which produces no H3K36me, and a *set2*Δ strain rescued with wild-type *SET2*. Western blot analysis of Set2 and H3K36me in the *set2* lysine point mutants revealed that these strains produced Set2 and H3K36me to wild-type levels (Figure 3.1B). Overall, these results established that individual sites of lysine methylation and acetylation on Set2 do not affect global Set2 or H3K36me levels.

set2 Lysine Point Mutants Do Not Affect Set2 Function in Certain Cellular Contexts

While global Set2 and H3K36me levels were unaffected by the *set2* lysine point mutants, we wanted to determine if the mutations affected Set2 function. To test this, we examined the *set2* lysine mutants under a variety of conditions. One phenotype we tested was the ability of *set2* lysine mutants to repress cryptic transcription, which is one of the most studied Set2 phenotypes. Without Set2 and H3K36me, cryptic transcripts arise from within gene bodies and can affect normal gene expression (McDaniel et al., 2017). To see if this occurred in the *set2* lysine mutants, we used the *FLO8* reporter, which has an internal cryptic initiation site (CIS) and the *HIS3* gene integrated out-of-frame downstream of the CIS. When transcription occurs from the canonical *FLO8* start site, *HIS3* is not transcribed and cells cannot grow on media lacking histidine. However, if transcription starts from the CIS, then *HIS3* is transcribed and cells can grow on media lacking histidine (Figure 3.2A) (Silva et al., 2012). Additionally, we examined the sensitivity of the *set2* lysine mutants to 6-azauracil (6AU, a transcriptional elongation inhibitor)

(Kizer et al., 2005). As shown previously, *set2*Δ cells were unable to repress cryptic transcription and resistant to 6AU (Figures 3.2B and 3.2C). In contrast, under all conditions tested, the *set2* lysine point mutants phenocopied wild-type Set2 (Figures 3.2B and 3.2C). Together, these results indicate that individual sites of lysine methylation and acetylation are not functionally relevant for cryptic transcription repression or transcriptional elongation.

Predicted Phosphorylation Sites are Found on Set2

In addition to the MUD-PIT analysis, we did additional mass spectrometry for phosphorylation sites. Through this analysis, three sites of serine phosphorylation were identified as the N-terminus of Set2, at residues S6, S8, and S10 (Figures 3.3A and 3.3C). We also took advantage of bioinformatic analysis through BioGRID and Eukaryotic Linear Motif Recourse (ELM), which curate information about protein PTMs and functional short linear motifs (Gouw et al., 2018; Oughtred et al., 2018). BioGRID predicted serine phosphorylation residues at S6, S8, and S10, in accordance with the mass spectrometry data, as well as phosphorylation at residue S522, which is found in the region between the WW and coiled-coiled domain (Figure 3.3B and 3.3C). ELM identified S357 in the AID as part of a Plk1 phosphorylation motif (DDDSLRH) and S726 at the C-terminus of Set2 as part of a Gsk3 phosphorylation motif (RMSSPPPS) (Figures 3.3B and 3.3C). We were particularly interested in residues S357 and S726 and their connection to cell-cycle genes since S6, S8, S10, and S522 were all identified in a screen for potential Cdk1 target sites, a major cell-cycle regulator (Holt et al., 2009). Overall, these results demonstrate that several serine residues on Set2 are phosphorylated and may have functionally important roles.

***set2* Serine Point Mutants Affect Set2 Protein Abundance, but not H3K36 Methylation**

We next wanted to examine how the loss of serine phosphorylation on Set2 would affect Set2 and H3K36me *in vivo*. Thus, we generated plasmids with a serine to alanine point mutation

for each of the residues previously identified and transformed them into a *set2*Δ strain. As controls, we included a *set2*Δ strain and *set2*Δ rescued with wild-type *SET2*. Interestingly, the set2-S6A and set2-S8A mutants had decreased Set2 protein levels compared to wild-type and set2-S10A had increased Set2 protein levels, but H3K36me was similar to wild-type in all three *set2* serine point mutants (Figure 3.3D). In contrast, *set2-S357A*, *set2-S522A*, and *set2-S726A* all had Set2 and H3K36me protein levels similar to wild-type (Figure 3.3D). Overall, these results demonstrate that the N-terminal serine residues and their predicted phosphorylation are functionally important for maintaining wild-type levels of Set2 protein in the cell.

set2 Serine Point Mutants Repress Cryptic Transcription at Reporter Loci but Have Transcriptional Elongation and DNA Double Strand Break Repair Defects

We next wanted to determine if the predicted sites of serine phosphorylation had functional significance, particularly the residues that when mutated to alanine affected Set2 protein levels. Similar to the *set2* lysine point mutants, we tested the *set2* serine point muants' ability to repress cryptic transcription at the *FLO8* locus and their sensitivity to 6AU and phleomycin (Jha and Strahl, 2014; Kizer et al., 2005; Silva et al., 2012). All *set2* serine point mutants were able to repress cryptic transcription at the *FLO8* locus, which is consistent with previous data that when wild-type levels of H3K36me are present, cells are able to repress cryptic transcription (Figures 3.4A and 3.4B). In contrast, all *set2* serine point mutants show some resistance to 6AU, displaying an intermediary phenotype between resistant *set2*Δ cells and sensitive wild-type cells (Figure 3.4C). Finally, all *set2* serine point mutants were sensitive to phleomycin, like *set2*Δ (Figure 3.4D). Overall, these results indicate that serine phosphorylation does not play a functional role in repressing cryptic transcription, but it is important for transcriptional elongation and DNA double strand break repair.

Discussion

Set2 is a histone methyltransferase with important transcription regulation functions. However, it was unknown if any PTMs were present on Set2 itself and if those modifications had functional significance. Through mass-spectrometry and bioinformatics analysis, we identified sites of lysine methylation and acetylation and serine phosphorylation on Set2. Further biochemical analysis revealed that individual lysine point mutants did not affect Set2 protein abundance, H3K36me, or some known functions of Set2. However, three serines at the N-terminus of Set2 showed functional importance in maintaining wild-type Set2 protein levels. Additionally, mutating predicted sites of serine phosphorylation across Set2 affected its transcriptional elongation function and repair DNA double-strand breaks. Overall, these data demonstrate the functional significance of PTMs on Set2.

While our data provide evidence for phospho-serine regulation dependent mechanisms of Set2, additional work is needed to elucidate the specific mechanism. Importantly, identifying the kinase will provide clues about the function of phosphorylation on Set2. A screen looking for Cdk1 (also called Cdc28) phosphorylation sites in budding yeast identified S6, S8, and S10 on Set2 as potential targets and another study demonstrated that Set2 protein abundance is highest in G2/M (Dronamraju et al., 2017; Holt et al., 2009). Thus, Cdk1 could phosphorylate Set2 during that time to regulate Set2 function. Additionally, the sensitivity of the *set2* serine mutants to phleomycin suggests a connection to the already established role of Set2 in DSB repair (Jha and Strahl, 2014). Interestingly, *set2-S6A* and *set2-S8A* were highly abundant and *set2-S10A* was lowly abundant compared to wild-type, but H3K36me levels were similar to wild-type in all three mutants. One possible explanation is that *set2-S6A* and *set2-S8A* might be hypoactive, while *set2-S10A* is hyperactive. Alternatively, Set2 serine phosphorylation could be unrelated to Set2 enzymatic activity and its main purpose could be to create a binding site for protein interaction. Future studies examining cell-cycle dependent functions of Set2

phosphorylation, enzymatic assays, and identifying potential protein interactions would provide insight into the role of serine phosphorylation on Set2.

While the individual *set2* lysine point mutants did not show any functional consequences in the biochemical and phenotypic assays tested, a combination of modifications on those sites could be important for Set2 regulation and function. Modeling the 3D structure or solving the crystal structure of Set2 would allow for a more throughout analysis of these residues and how their conformation in 3D space might suggest logical combinations to test. Additionally, identifying the enzymes responsible for methylation and acetylation of Set2 could provide clues as to their functionality. One possibility is that Set2 automethylates itself to regulate its function. Recently, an auto-inhibitory domain (AID) was identified in Set2 and demonstrated to function along with the SRI domain to control Set2 substrate specificity (Wang et al., 2015). Automethylation could work together with the AID and SRI to ensure proper H3K36me throughout the genome, similar to how Gcn5 acetylation of Rsc4 controls H3K14ac (VanDemark et al., 2007). Alternatively, other enzymes could methylate or acetylate Set2 to modulate its function. Examining the Set2 protein sequence for sequences that are similar to motifs that other methyltransferases and acetyltransferases target would suggest enzymes for further investigation. In addition to potentially modulating its enzymatic activity, methylation or acetylation of Set2 could serve as protein binding sites or stabilize the Set2 protein.

Emerging data demonstrate that PTMs on chromatin-modifying enzymes are functionally important, but this evidence was lacking for Set2 (Biggar and Li, 2015; Narita et al., 2019; Ubersax and Ferrell, 2007). We identified and examined sites of methylation, acetylation, and phosphorylation and found functional significance for three sites of serine phosphorylation on Set2. Our findings establish the groundwork for elucidating the mechanism establishing these PTMs and how they affect cellular processes.

Materials and Methods

Yeast Strains and Plasmids

Yeast strains were created using the PCR Toolbox (Janke et al., 2004). Yeast strains used in this study are listed in Table 3.2. All plasmids were created using site-directed mutagenesis (Agilent) with the primers in Table 3.3. Plasmids used in this study are listed in Table 3.4.

Mass Spectrometry

TAP tagged full-length Set2 and set2-1-618 were purified from BY4741 cells and analyzed via MudPIT for methylation and acetylation (Washburn et al., 2001). To identify phosphorylated residues, 32P-labeled Set2 samples were trypsinized, purified by LC-MS, and characterized by MALDI-TOF spectrometry (Ficarro et al., 2002).

Database Searching

SET2 (YJL168C) was queried in BioGRID version 3.5 to identify phospho-serine sites (Oughtred et al., 2018). SET2_YEAST (UniProt P46995) was queried in ELM and S357 was identified as part of a polo-like kinase (Plk) recognition motif (DDDSLRH) and S726 was identified as part of a Gsk3 recognition motif (SASTRMSS) (Gouw et al., 2018).

Western Blotting

All strains were grown to saturation in SC-Ura before being diluted to an OD_{600} of 0.1 and grown to OD_{600} ~1 in SC-Ura at 30°C. Ten ODs of cells were collected and Western blotting was performed after protein extraction by SUMEB lysis (Jha and Strahl, 2014). Lysates were separated by SDS-PAGE, transferred to PVDF membrane, and probed overnight at 4°C with Set2 (in-house), G6PDH (Sigma A9521-1VL), H3K36me1 (abcam 9048), H2K36me2 (Active Motif 39255), H3K36me3 (abcam 9050), or H3 C-terminal (EpiCypher 13-0001). Membranes were washed in TBS (Tris-buffered saline)-Tween (50 mM Tris-HCl, 150 mM NaCl, and 0.5% Tween 20). Membranes were incubated with HRP-conjugated anti-rabbit (GE Healthcare NA934V; 1:10,000) antibody and probed with ECL Prime (GE Healthcare).

Spotting Assays

Yeast strains were grown at 30°C to saturation in SC-Ura and diluted to an OD_{600} of 0.5 prior to spotting 5-fold serial dilutions on plates at 30°C for 2-4 days. All experiments were performed in technical triplicate.

Tables

Table 3.1: Lysine Residues on Set2 are Methylated and Acetylated

Residue	Post-Translational Modification	Domain Location
K117	Dimethylation	AWS
K126	Dimethylation	SET
K228	Acetylation/Trimethylation	SET
K247	Acetylation/Trimethylation	PS
K376	Dimethylation	AID
K376	Acetylation/Trimethylation	AID
K438	Acetylation/Trimethylation	AID
K447	Dimethylation	AID
K530	Acetylation/Trimethylation	
K584	Acetylation/Trimethylation	CC

Summary of residues identified by MudPIT analysis to be methylated and/or acetylated on Set2 and their domain location.

Table 3.2: List of Yeast Strains and Genotypes

Name	Genotype	Reference
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Open Biosystems
BY4742	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	Open Biosystems
BY4741 Set2-TAP	MATa leu2Δ0 met15Δ0 ura3Δ0 SET2-TAP::HIS	Dharmacon
BY4741 set2-1-618-TAP	MATa leu2Δ0 met15Δ0 ura3Δ0 set2-1-618-TAP::HIS	This study
Set2D	MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 set2::KanMX6	Xiao et al., 2003
KLY78	MATalpha ura3-52 leu2Δ1 trp1Δ63 his3Δ200 KanMX6::GAL1pr-flo8::*HIS3* lys2-128δ	Nourani, Robert, & Winston, 2006
YSM138	MATalpha ura3-52 leu2Δ1 trp1Δ63 his3Δ200 KanMX6::GAL1pr-flo8::*HIS3* lys2-128δ set2::NatMX	This study

Table 3.3: List of Primers

Name	Sequence	Target	Source
K117AF	GGCTATCGGAGCGTACTGCGCCTTTTGAAATCTCTGATTTTG	Forward primer *set2-K117A* for site directed mutagenesis	This study
K117AR	CAAAATCAGAGATTTCAAAAGGCGCAGTACGCTCCGATAGCC	Reverse primer *set2-K117A* for site directed mutagenesis	This study
Set2K126AF	GAAGCAGTACGCTCCGATAGCCATATTTGCAACAAAGCATAAAGGTT	Forward primer *set2-K126A for* site directed mutagenesis	This study
Set2K126AR	AACCTTTATGCTTTGTTGCAAATATGGCTATCGGAGCGTACTGCTTC	Reverse primer *set2-K126A for* site directed mutagenesis	This study
Set2K228AF	AAGTGATTTCCTCACCTGCTAAAATTTTTCTTTGAGCAAATATTCCCATGCG	Forward primer *set2-K228A for* site directed mutagenesis	This study
Set2K228AR	CGCATGGGAATATTTGCTCAAAGAAAAATTTTAGCAGGTGAGGAAATCACT	Reverse primer *set2-K228A for* site directed mutagenesis	This study
Set2K247AF	TATGGTGCTCAAGCTCAGGCATGCTACTGTGAGGAGCC	Forward primer *set2-K247A for* site directed mutagenesis	This study
Set2K247AR	GGCTCCTCACAGTAGCATGCCTGAGCTTGAGCACCATA	Reverse primer *set2-K247A for* site directed mutagenesis	This study
Set2K438AF	ACTCTCCCTGCAAAGTTTCCATTTTTAGCGCCAAACTGCGATG	Forward primer *set2-K438A for* site directed mutagenesis	This study
Set2K438AR	CATCGCAGTTTGGCGCTAAAAATGGAAACTTTGCAGGGAGAGT	Reverse primer *set2-K438A for* site directed mutagenesis Delitto Perfetto	This study
Set2K530AF	AAGCGCCAGGCGTTGCTGGTATTCCACTGCTACTCG	Forward primer *set2-K530A for* site directed mutagenesis	This study
Set2K530AR	CGAGTAGCAGTGGAATACCAGCAACGCCTGGCGCTT	Reverse primer *set2-K530A for* site directed mutagenesis	This study
Set2K584AF	CAAAAGCTAAAGCTTGAGAATGAAAGAGCAAGCGTTTTGGAGGATATTATAGC	Forward primer *set2-K584A for* site directed mutagenesis	This study
Set2K584AR	GCTATAATATCCTCCAAAACGCTTGCTCTTTCATTCTCAAGCTTTAGCTTTTG	Reverse primer *set2-K584A for* site directed mutagenesis	This study
S726AFor	CATCAACAAGGATGTCTGCTCCTCCACCTTCAACA	Forward primer *set2-S726A for* site directed mutagenesis	This study
S726ARev	TGTTGAAGGTGGAGGAGCAGACATC	Reverse primer *set2-*	This study

	CTTGTTGATG	*S726A for* site directed mutagenesis	
K376AF	CTGAGGTTGTTCCGTGATAAACAATG CAAGCATTTTGCTAAAACAGGTGTAA	Forward primer *set2-K376A for* site directed mutagenesis	This study
K376AR	TTACACCTGTTTTAGCAAAATGCTTG CATTGTTTATCACGGAACAACCTCAG	Reverse primer *set2-K376A for* site directed mutagenesis	This study
S357A_For	GATAGCCTGATGACGAAGAGCGTCA TCATCGATAGTAAAAA	Forward primer *set2-S357A for* site directed mutagenesis	This study
S357A_Rev	TTTTTACTATCGATGATGACGCTCTTC GTCATCAGGCTATC	Reverse primer *set2-S357A for* site directed mutagenesis	This study
S522A_For	ACTGCTACTCGGAGCATTTACTTGTG TATTTGACCTTGAAC	Forward primer *set2-S522A for* site directed mutagenesis	This study
S522A_Rev	GTTCAAGGTCAAATACACAAGTAAAT GCTCCGAGTAGCAGT	Reverse primer *set2-S522A for* site directed mutagenesis	This study
S6A_For	CGACGCACTCACAGCTTGGTTCTTCG ACATGATCACGT	Forward primer *set2-S6A for* site directed mutagenesis	This study
S6A_Rev	ACGTGATCATGTCGAAGAACCAAGCT GTGAGTGCGTCG	Reverse primer *set2-S6A for* site directed mutagenesis	This study
S8A_For	TTTTTCATCTTCCGACGCAGCCACAC TTTGGTTCTTCGAC	Forward primer *set2-S8A for* site directed mutagenesis	This study
S8A_Rev	GTCGAAGAACCAAAGTGTGGCTGCG TCGGAAGATGAAAAA	Reverse primer *set2-S8A for* site directed mutagenesis	This study
JDS10A_F	ATTTCTTTTTCATCTTCCGCCGCACTC ACACTTTGGTTC	Forward primer *set2-S10A for* site directed mutagenesis	This study
JDS10A_R	GAACCAAAGTGTGAGTGCGGCGGAA GATGAAAAAGAAAT	Reverse primer *set2-S10A for* site directed mutagenesis	This study

Table 3.4: List of Plasmids

Name	Features	Description	Reference
pRS416	CEN6_ARS4, URA3, Amp	Empty Vector	Sikorski and Hieter, 1989
pRS416-Set2-3xFlag	CEN6_ARS4, URA3, Amp	*SET2*-3xFlag	Du et al., 2008
pRS416 Set2 (K117A)	CEN6_ARS4, URA3, Amp	*set2-K117A*-3xFlag	This study
pRS416 Set2 (K126A)	CEN6_ARS4, URA3, Amp	*set2-K126A*-3xFlag	This study
pRS416 Set2 (K228A)	CEN6_ARS4, URA3, Amp	*set2-K228A*-3xFlag	This study
pRS416 Set2 (K247A)	CEN6_ARS4, URA3, Amp	*set2-K247A*-3xFlag	This study
pRS416 Set2 (K376A)	CEN6_ARS4, URA3, Amp	*set2-K376A*-3xFlag	This study
pRS416 Set2 (K438A)	CEN6_ARS4, URA3, Amp	*set2-K438A*-3xFlag	This study
pRS416 Set2 (K447A)	CEN6_ARS4, URA3, Amp	*set2-K447A*-3xFlag	This study
pRS416 Set2 (K530A)	CEN6_ARS4, URA3, Amp	*set2-K530A*-3xFlag	This study
pRS416 Set2 (K584A)	CEN6_ARS4, URA3, Amp	*set2-K584A*-3xFlag	This study
pRS416 Set2 (S6A)	CEN6_ARS4, URA3, Amp	*set2-S6A*-3xFlag	This study
pRS416 Set2 (S8A)	CEN6_ARS4, URA3, Amp	*set2-S8A*-3xFlag	This study
pRS416 Set2 (S10A)	CEN6_ARS4, URA3, Amp	*set2-S10A*-3xFlag	This study
pRS416 Set2 (S357A)	CEN6_ARS4, URA3, Amp	*set2-S357A*-3xFlag	This study
pRS416 Set2 (S522A)	CEN6_ARS4, URA3, Amp	*set2-S522A*-3xFlag	This study
pRS416 Set2 (S726A)	CEN6_ARS4, URA3, Amp	*set2-S726A*-3xFlag	This study

Figures

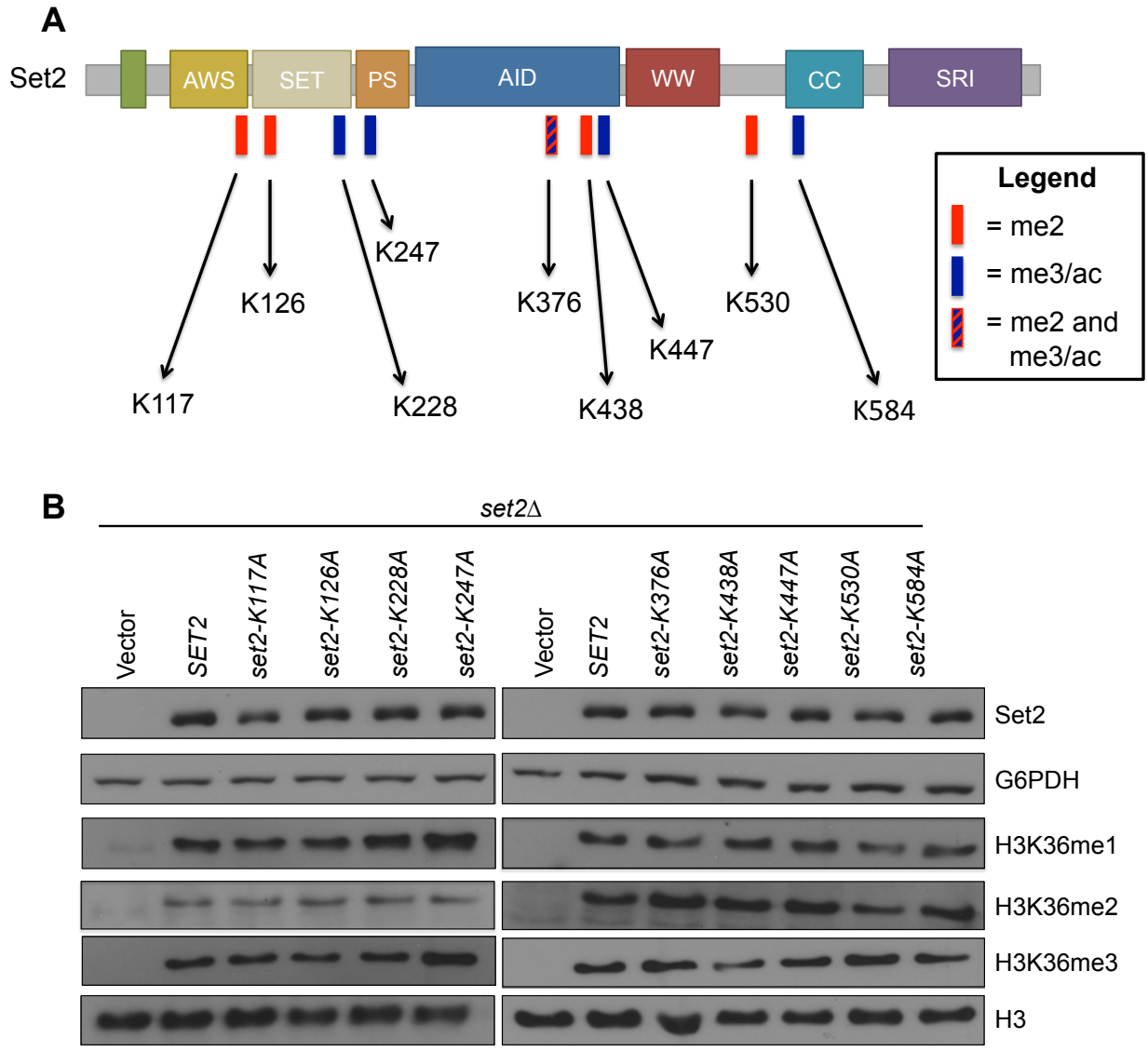


Figure 3.1: *set2* Lysine Point Mutants Do Not Affect Set2 Protein Abundance or H3K36 Methylation

A. Domain map of Set2 with the sites of methylation and/or acetylation switch highlighted. The catalytic region of Set2 is composed of the associated with SET (AWS), catalytic SET domain, and post-SET (PS). The autoinhibitory domain (AID) regulates catalytic activity. The C-terminus of Set2 has protein-protein interaction domains, like the coiled-coiled (CC), WW, and Set2-Rpb1 (SRI) protein-protein interaction domain. Sites in red were detected as dimethylated, sites in blue were detected as trimethylated or acetylated, and the site in blue and red stripes was detected as dimethylated and trimethylated or acetylated. B. Western blots of yeast strains transformed with indicated plasmids and probed with Set2 and different H3K36 methylation antibodies. G6PDH and H3 served as loading controls.

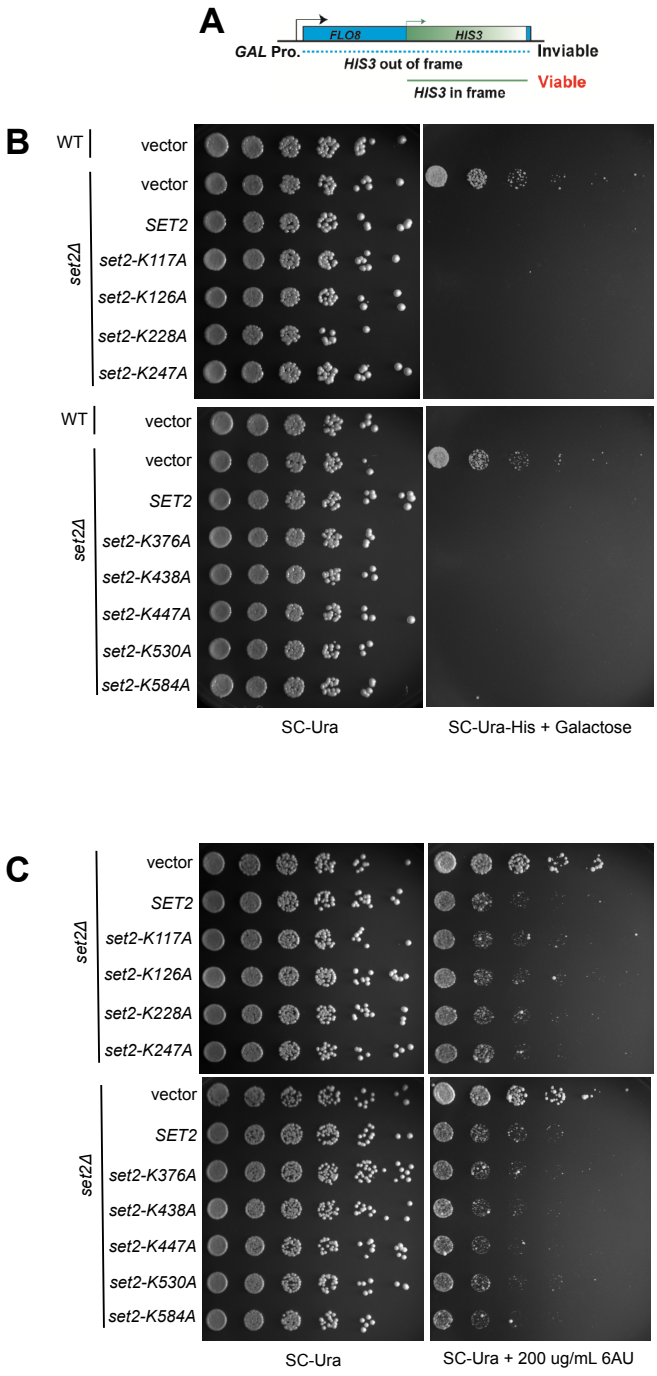
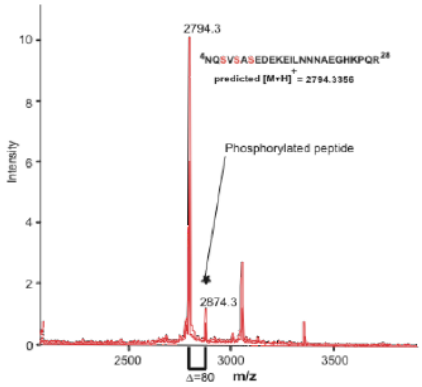


Figure 3.2: *set2* Lysine Point Mutatns Do Not Affect Set2 Function in Certain Cellular Contexts

A. Schematic of *FLO8 -HIS3* fusion gene reporter to detect cryptic transcription. B. Five-fold serial dilutions of indicated wild-type, *set2*Δ, *SET2* rescue, and *set2* lysine mutant strains plated on SC-Ura or SC-Ura-His with 2% galactose. C. Five-fold serial dilutions of *set2*Δ, *SET2* rescue, and *set2* lysine mutant strains plated on SC-Ura or SC-Ura with 6-AU (200 µg/mL).
Figure 3.3: Predicted Phosphorylation Sites are Found on Set2 and *set2* Serine Point Mutants Affect Set2 Protein Abundance, but not H3K36 Methylation

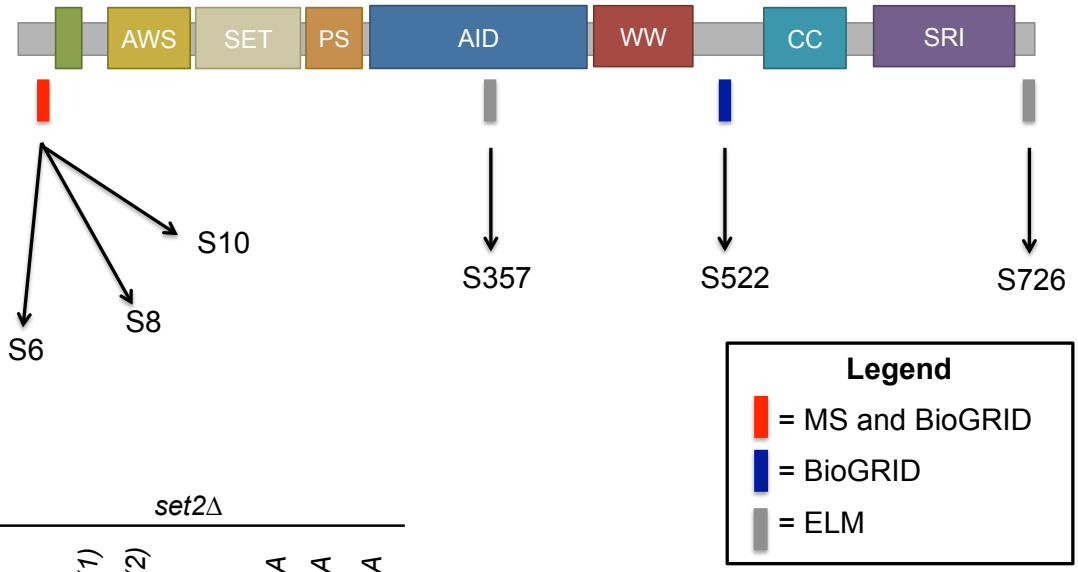
A



B

Residue	Source	Domain Location
S6	MS and BioGRID	
S8	MS and BioGRID	
S10	MS and BioGRID	
S357	ELM	AID
S522	BioGRID	
S726	ELM	

C



D

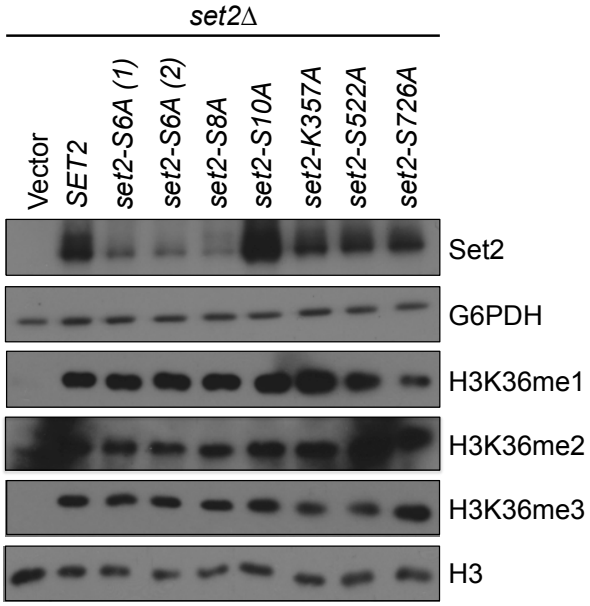


Figure 3.3: Predicted Phosphorylation Sites are Found on Set2 and *set2* Serine Point Mutants Affect Set2 Protein Abundance, but not H3K36 Methylation

A. Trypsinized samples of ^{32}P-labeled Set2 were purified by LC-MS and characterized by MALDI-TOF spectrometry. One protein in the N-terminus of Set2 was found to be phosphorylated ∆=80amu). B. Summary of residues identified by MS, BioGRID, and ELM to be phosphorylated on Set2 and their domain location. C. Domain map of Set2 with the sites of methylation and/or acetylation switch highlighted. The catalytic region of Set2 is composed of the associated with SET (AWS), catalytic SET domain, and post-SET (PS). The autoinhibitory domain (AID) regulates catalytic activity. The C-terminus of Set2 has protein-protein interaction domains, like the coiled-coiled (CC), WW, and Set2-Rpb1 (SRI) protein-protein interaction domain. Sites in red were detected by MS and BioGRID, the site in blue was detected by BioGRID, and the sites in gray were detected by ELM. D. Western blots of yeast strains transformed with indicated plasmids and probed with Set2 and different H3K36 methylation antibodies. G6PDH and H3 served as loading controls.

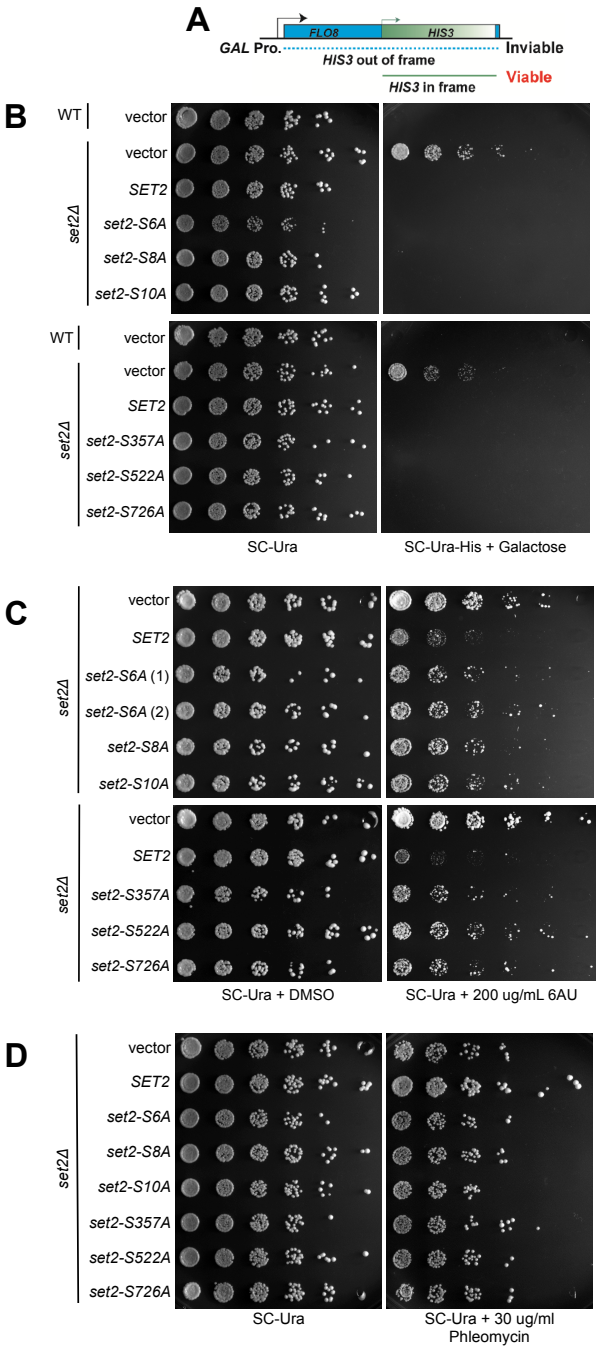


Figure 3.4: *set2* Serine Point Mutants Repress Cryptic Transcription at Reporter Loci and Have Transcriptional Elongation and DNA Double Strand Repair Defects

A. Schematic of *FLO8 -HIS3* fusion gene reporter to detect cryptic transcription. B. Five-fold serial dilutions of indicated wild-type, *set2*Δ, *SET2* rescue, and *set2* serine mutant strains plated on SC-Ura or SC-Ura-His with 2% galactose. C. Five-fold serial dilutions of *set2*Δ, *SET2* rescue, and *set2* serine mutant strains plated on SC-Ura or SC-Ura with 6-AU (200 µg/mL). D. Five-fold serial dilutions of set2Δ, SET2 rescue, and *set2* serine mutant strains plated on SC-Ura or SC-Ura with phleomycin (30 µg/mL).

CHAPTER 4 – CONCLUSIONS AND FUTURE DIRECTIONS

H3K36me States Work Redundantly to Repress Cryptic Transcription During Nutrient Stress

One of the major findings of this work is that H3K36me1/2 and H3K36me3 are redundant in their ability to repress cryptic transcription during nutrient stress. Previous work examining the functions of the different H3K36me states in cryptic transcription focused on cells completely lacking H3K36me or those that only have H3K36me1 or H3K36me1/2. In that context, cells with H3K36me1/2 were still able to repress cryptic transcription and H3K36me3 was considered dispensable (Hacker et al., 2016; Li et al., 2009; Youdell et al., 2008). However, by utilizing the Phe/Tyr switch in the Set2 SET domain, we created a mutant (*set2-Y149F*) that predominately performed H3K36me3 *in vivo*. Thus, the *set2-Y149F* mutant allowed us to study the contributions of H3K36me3 alone. Combined with *set2-F234Y* and *set2-H199L* , which catalyze H3K36me1/2 and H3K36me1, respectively, we comprehensively examined the functions of the different H3K36me states in various cellular processes, particular repressing cryptic transcription after nutrient stress.

The redundancy of H3K36me1/2 and H3K26me3 may provide flexibility to the cell when responding to nutrient stress. During a time when the cell must rapidly change its transcriptional programming, it would be advantageous for multiple histone modifications to have shared functions. The shared functionality provides more opportunities for other critical chromatin-modifying enzymes to perform their normal activity. Previous *in vitro* studies demonstrated that Eaf3 and Ioc4 bind H3K36me2 and H3K36me3 peptides (Carrozza et al., 2005; Joshi and Struhl, 2005; Li et al., 2009; Maltby et al., 2012). Structural analysis of the Eaf3 chromodomain

shows that it is part of a group of chromodomains that lack a carboxylate group, thus allowing it to bind to mono-, di-, and trimethylated lysines (Xu et al., 2008). While there is no detailed structural data for the PWWP domain of Ioc4, other PWWP domains have been shown to bind H3K36me2 and H3K36me3 (Rona et al., 2016). The *in vivo* data presented in this work aligns with the previous *in vitro* results and structural analyses, thus supporting a model where Eaf3 and Ioc4 can bind H3K36me1/2 and H3K36me3 to carry out their deacetylating and nucleosome remodeling activities, respectively. Similarly, H3K36me2 and H3K36me3 peptides were poor binding substrates for Asf1 (Venkatesh et al., 2012). While Asf1 does not make any contacts with H3K36 when it forms a heterotrimeric complex with the H3/H4 heterodimer during nucleosome exchange, other factors may influence its inability to bind nucleosomes containing H3K36me (English et al., 2006). Overall, the redundancy of H3K36me1/2 and H3K36me3 in suppressing cryptic transcription during nutrient stress prevents aberrant gene expression and allows the cell to successfully respond to its environment.

While wild-type, *set2-Y149F* , and *set2-F234Y* strains are able to similarly repress cryptic transcription in response to nutrient stress, overall changes to their chromatin architecture remain unknown. ChIP-qPCR presented in this work suggests that *set2-Y149F* has similar H3K36me3 distribution as wild-type and *set2-F234Y* generally distributes H3K36me2 similarly to wild-type, but future ChIP-seq studies examining the distribution of the H3K36me states before and after nutrient stress will be needed to provide further insight into their redundancy. Furthermore, *set2-F234Y* deposits significantly more H3K36me1 than wild-type and determining how this extra monomethylation is deposited genome wide could provide insight into the function of H3K36me1. Additionally, examining other histone PTMs and their correlation with H3K36me would provide a better understanding of how histone PTMs contribute to repressing cryptic transcription. In particular, examining the genome-wide distribution of H3K56ac and H3K27ac. As detected by ChIP-qPCR, wild-type, *set2-Y149F* , and *set2-F234Y* generally had similar levels of these acyl marks near CISs, but this remains to be tested genome-wide. Also,

examining promoter-associated histone PTMs, such as H3K4mes, H3K18ac, and Htz1 and their distribution near CISs would provide insight into what promoter features, if any, contribute to transcriptional initiation (Rando and Winston, 2012; Sadeh et al., 2016).

Interestingly, Western blot analysis indicated that before nutrient stress, there is approximately 50% less H3K36me3 present in *set2-Y149F* than wild-type. However, after nutrient stress, there are similar levels of H3K36me3 presented in *set2-Y149F* and wild-type. Quantitative mass spectrometry measuring the total levels of each H3K36me state in the different mutants would allow for improved characterization of the mutants and insight into how much H3K36me is required to suppress cryptic transcription (Noberini et al., 2016; Zheng et al., 2016). Likewise, directly assessing chromatin accessibility in the wild-type and *set2* mutants would show how nucleosome positioning affects cryptic transcription. Previous work in ccRCC tumor samples demonstrated that regions with increased chromatin accessibility in SETD2 mutants overlapped with regions where H3K36me3 was normally found (Simon et al., 2014). Additionally, studies in budding yeast showed that cryptic transcripts often arise from NFRs (Neil et al., 2009; Xu et al., 2009). Using FAIRE-seq or MNase-seq and examining nucleosome occupancy in wild-type and *set2* mutants would further our understanding of H3K36me, nucleosome occupancy, and cryptic transcription.

Further lines of inquiry that remain unaddressed are if cryptic transcription occurs during other forms of stress, if the same CISs are affected, and if the same mechanisms control repression. Since all forms of stress require the cell to rapidly adjust its transcriptional programming, and thus their chromatin architecture, it stands to reason that cryptic transcription occurs during heat stress, oxidative stress, and osmotic stress, amongst others. Intriguingly, another analysis of cryptic transcription during nutrient stress saw that many genes with CISs were not connected to nutrient response (McDaniel et al., 2017). Thus, similar CISs could be affected during different types of stress. However, certain CISs may be more prone to utilization depending on how the chromatin architecture is disrupted. Similar to how different classes of

cryptic transcripts have been established based on their sensitivity to degradation pathways, certain CISs may be associated with defects in particular chromatin modifying enzymes (van Dijk et al., 2011; Thompson and Parker, 2007; Wyers et al., 2005; Xu et al., 2009). One class of cryptic transcripts that arise in *set2*Δ cells under normal growth conditions has been classified as Set2-repressed antisense transcripts, or SRATs (Venkatesh et al., 2016). Other distinct classes of cryptic transcripts may be identified in the future in spt6 mutants, FACT complex mutants, amongst others. Establishing the interplay between stress and chromatin-modifying enzymes and how that affects transcription will provide insight into important regulatory mechanisms.

H3K36me3 Has a Unique Function Related to Bur1 and Spt16

While H3K36me1/2 and H3K36me3 were redundant in repressing cryptic transcription during nutrient stress, H3K36me3 had a unique function related to Bur1 and Spt16 that H3K36me1/2 could not compensate. While Bur1 is an essential kinase and contributes to serine 2 phosphorylation on the RNAPII CTD, *bur1*Δ *set2*Δ cells can survive (Chu et al., 2006). The observation that cells lacking both Bur1 and Set2 can survive led to the idea that Bur1 and Set2 have opposing functions in transcription elongation, but there is still no clear understanding of their opposing roles. Intriguingly, previous studies demonstrate that bur1 mutants lack H3K36me3 (Chu et al., 2006). In this work, *bur1*Δ *set2-Y149F* , which predominately performs H3K36me3, resulted in cells being unable to survive, phenocopying *bur1*Δ cells. Perhaps H3K36me3 has a role that is antagonistic to Bur1 activity. However, the H3K36me3 status of these cells was not confirmed. Another possibility is that *bur1*Δ *set2-Y149F* cells do not have any H3K36me3 and therefore no H3K36me at all. Such a finding would indicate that *bur1*Δ *set2-Y149F* cells have the same H3K36me status as *bur1*Δ *set2*Δ, yet the strains have different phenotypes. In that case, something other than H3K36me could be contributing to the observed

phenotypic differences. Future work characterizing the H3K36me status of the strains harboring *bur1*Δ combined with different *set2* mutants will be critical in furthering our understanding of the relationship between Bur1 and H3K36me.

Spt16 is part of the FACT complex and is responsible for nucleosome reorganization (Belotserkovskaya et al., 2003). Previous *in vitro* data suggests that H3K36me3 peptides are poor binding substrates for Spt16, while Spt16 can bind to unmodified H3 and H3K36me2 peptides (Venkatesh et al., 2012). In contrast, *in vivo* results show that *set2*Δ cells have similar Spt16 occupancy as wild-type at the *ARG1* and *ADH1* loci (Pathak et al., 2018). Thus, it remains unclear how H3K36me3 specifically affects Spt16 binding and function. Future work examining how Spt16 occupancy is affected genome-wide in the presence or absence of different forms of H3K36me will be critical to elucidating the relationship between FACT and H3K36me. Additionally, histone exchange assays with the *set2* mutants presented in this study will reveal how the different H3K36me forms affect histone exchange. The initial work that connected Set2 to histone exchange was done with wild-type and *set2*Δ cells, thus limiting our understanding of H3K36me and its involvement in nucleosome exchange (Venkatesh et al., 2012). Also, establishing if Set2 and Spt16 or other members of the FACT complex directly interact would provide further insight into their relationship. Using photocrosslinking, Cucinotta et al., 2019 demonstrates that Spt16 specifically contacts the acidic patch on H2A. A similar technique could be used for the residues near H3K36 to determine how that region interacts with other chromatin-modifying enzymes like Spt16. Overall, a variety of data connects Set2 and Spt16 function and a mechanistic understanding of their relationship will provide insight into nucleosome reorganization and its influence on chromatin structure.

Set2 Post-Translational Modifications Affect Its Protein Abundance, Transcriptional Elongation Role, and Ability to Repair DNA Double Stranded Break

Another major contribution of this work is the identification of PTMs on Set2 and how they impact Set2 abundance and function. While the sites of lysine methylation and acetylation did not show individual functional significance, three sites of serine phosphorylation at the N-terminus of Set2 were important for maintaining wild-type Set2 protein levels. Interestingly, despite dramatic changes in Set2 protein abundance in *set2-S6A*, *set2-S8A*, and *set2-S10A*, all mutants had similar H3K36me levels to wild-type. While none of the putatively phosphorylated serine residues appear to play a role in repressing cryptic transcription, they do show functional importance during transcriptional elongation and DNA DSB repair.

The identification of phospho-serine sites on Set2 provides exciting potential for uncovering key regulatory mechanisms of the Set2 protein. While much is known regarding the role of Set2 in transcriptional elongation, there is little understanding about how Set2 is recruited to chromatin or roles outside of histone lysine methylation. Since the H3K36me levels of all *set2* serine mutants are similar to wild-type, it appears that Set2 recruitment to chromatin is unaffected by the loss of serine phosphorylation. However, the wild-type levels of H3K36me despite the decreased Set2 protein levels in *set2-S6A* and *set2-S8A* and increased Set2 protein levels in *set2-S10A* suggests that the catalytic activity of Set2 is affected. Future work measuring the enzymatic activity of *set2-S6A*, *set2-S8A*, and *set2-S10A* would determine if these mutants are hypo- or hyperactive. If the enzymatic activity of the set2 N-terminal serine mutants is altered, then steady state levels of H3K36me appear the same as wild-type, but the temporal regulation could be altered, thus contributing to the transcriptional elongation defects observed on the 6AU assays. Using a recently developed light-activated nuclear shuttle (LANS) system, Set2 protein could be shuttled from the cytoplasm into the nucleus and the changes in H3K36me observed over time (Lerner et al., 2018). Additionally, the N-terminal serines were initially identified as putative phosphorylation sites in a screen for Cdk1 (also called Cdc28)

targets (Holt et al., 2009). Along with the established role for Set2 in cell cycle regulation, the phosphorylation sites could be critical for recruiting or binding to specific cell cycle proteins that help stabilize Set2 protein levels (Dronamraju et al., 2017). Finally, the phleomycin sensitivity detected in all the *set2* serine mutants points to a function in DSB repair (Jha and Strahl, 2014). Overall, the serine phosphorylation on Set2 could be connected to several known roles of Set2 and contribute to previously unknown regulatory mechanisms.

The single *set2* lysine point mutants did not have any functional significance in regulating Set2 or H3K36me levels or in the phenotypic assays tested, but there is still more to explore in regards to their functionality. One possibility is that the modifications work in combination rather than alone. However, a 3D model or crystal structure of Set2 would be necessary to determine logical combinations to test based on their conformation in 3D space. Another intriguing idea is that Set2 automethylates itself. Automethylation, combined with the recently discovered AID and its relationships to the SET and SRI domains, could be important for protein stability and enzymatic regulation (Wang et al., 2015). One challenge in testing for Set2 automethylation is the lack of a true Set2 catalytic dead mutant *in vivo*. While *set2-H199L* lacks H3K36me3 and H3K36me2, it does still perform low levels of H3K36me1, thus making it an insufficient negative control. Continued mutational analysis to find a true Set2 catalytic dead mutant will be the first step towards identifying a possible automethylation mechanism. In addition to automethylation, other proteins could methylate or acetylate Set2. While challenging to identify them, sequence analysis for motifs of known methyltransferases and acetyltransferases could provide insight into potential modifiers. Finally, the sites of methylation and acetylation could serve as binding sites for other proteins. Identifying such an interaction could provide further insight into the role of Set2 in transcription, DNA DSB repair, or its other functions. Overall, lysine methylation and acetylation provides intriguing possibilities to test and improve our understanding of Set2 regulation and function.

Final Thoughts

Substantial efforts have been made to elucidate the molecular mechanisms underpinning the diversity of the natural world. From Mendel to today, we have collectively amassed a great body of knowledge describing how cells store, utilize, and repair their genetic material. While some of these processes are different across organisms, many of the fundamental principles are conserved. Thus, I hope that the work presented here will provide the basis for future discoveries across species. Nevertheless, the findings of this dissertation have contributed, even in the smallest of ways, toward better understanding how genetic material is accessed and utilized. It is only through this incremental increase in knowledge that we can fully appreciate the diversity of the natural world and the complexity of life.

APPENDIX A – STRUCTURE/FUNCTION ANALYSIS OF RECURRENT MUTATIONS IN SETD2 REVEALS A CRITICAL AND CONSERVED ROLE FOR A SET DOMAIN RESIDUE IN MAINTAINING PROTEIN STABILITY AND H3K36 TRIMETHYLATION[1]

Summary

The yeast Set2 histone methyltransferase is a critical enzyme that plays a number of key roles in gene transcription and DNA repair. Recently, the human homologue, SETD2, was found to be recurrently mutated in a significant percentage of renal cell carcinomas (RCCs), raising the possibility that the activity of SETD2 is tumor suppressive. Using budding yeast and human cell line model systems, we examined the functional significance of two evolutionarily conserved residues in SETD2 that are recurrently mutated in human cancers. While one of these mutations (R2510H), located in the Set2 Rpb1 interaction (SRI) domain, did not result in an observable defect in SETD2 enzymatic function, a second mutation in the catalytic domain of this enzyme (R1625C) resulted in a complete loss of H3K36me3. This mutant showed unchanged thermal stability as compared to the wild type protein, but diminished binding to the histone H3 tail. Surprisingly, mutation of the conserved residue in Set2 (R195C) similarly resulted in a complete loss of H3K36me3, but did not affect H3K36me2 or functions associated with H3K36me2 in yeast. Collectively, these data imply a critical role for R1625 in maintaining the protein interaction with H3 and specific H3K36me3 function of this enzyme, which is conserved from yeast to humans. They also may provide a refined biochemical explanation for how H3K36me3 loss leads to genomic instability and cancer.

[1]Portions of this chapter were adapted from Hacker, K.E., Fahey, C.C., Shinsky, S.A., Chiang, Y.J., DiFiore, J.V., Jha, D.K., Vo, A.H., Shavit, J.A., Davis, I.J., Strahl, B.D., Rathmell, W.K. (2016). Structure/Function Analysis of Recurrent Mutations in SETD2 Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and H3K36 Trimethylation. *Journal of Biological Chemistry* 291, 21283-21295.

Introduction

Cancer is increasingly characterized by alterations in chromatin-modifying enzymes (Cancer Genome Atlas Research Network, 2013). SETD2, a non-redundant histone H3 lysine 36 (H3K36) methyltransferase, has been found to be mutated in a growing list of tumor types, most notably in clear cell renal cell carcinoma (ccRCC), but also in high-grade gliomas, breast cancer, bladder cancer and acute lymphoblastic leukemia (ALL) (Cancer Genome Atlas Research Network, 2013, 2014; Dalgliesh et al., 2010; Duns et al., 2010; Edmunds et al., 2008; Fontebasso et al., 2013; Mar et al., 2014; Al Sarakbi et al., 2009; Zhang et al., 2012; Zhu et al., 2014). Recent studies exploring intratumor heterogeneity in ccRCC identified distinct mutations in SETD2 from spatially distinct subsections of an individual tumor, suggesting that mutation of SETD2 is a critical and selected event in ccRCC cancer progression (Gerlinger et al., 2012). Mutations in SETD2 are predominantly inactivating, such as early nonsense or frameshift mutations, which lead to non-functional protein and global loss of H3K36me3 (Dalgliesh et al., 2010; Gerlinger et al., 2012; Simon et al., 2014). Missense mutations tend to cluster in two domains: the SET domain, which catalyzes H3K36 trimethylation (H3K36me3, and the SRI domain, which mediates the interaction between SETD2 and the hyperphosphorylated form of RNA Polymerase II (RNAPII) (Cancer Genome Atlas Research Network, 2013; Dalgliesh et al., 2010; Simon et al., 2014; Sun et al., 2005; Wagner and Carpenter, 2012).

SETD2, and its yeast counterpart, Set2, both associate with RNAPII in a co-transcriptional manner (Kizer et al., 2005; Sun et al., 2005; Yoh et al., 2008). In yeast, Set2 mediates all H3K36 methylation states (H3K36me1/me2/me3) and regulates the recruitment of chromatin-remodeling enzymes (Isw1b) and a histone deacetylase (Rpd3) that functions to keep gene bodies deacetylated, thereby maintaining a more compact chromatin structure that is more resistant to inappropriate and bi-directional transcription (Carrozza et al., 2005; Li et al., 2009; Lickwar et al., 2009; Quan and Hartzog, 2009; Smolle et al., 2012). The Set2/SETD2 pathway is also important for DNA repair in both yeast and humans, as well as for proper mRNA splicing

(Aymard et al., 2014; Carvalho et al., 2014; Daugaard et al., 2012; Fnu et al., 2011; Kanu et al., 2015; Luco et al., 2010; Pfister et al., 2014; Simon et al., 2014; Sorenson et al., 2016). Although yeast Set2 can mediate all forms of H3K36 methylation, SETD2 only trimethylates H3K36 (Sun et al., 2005). Other methyltransferses (e.g., NSD2 and ASH1L) mediate mono- and dimethylation, indicating an increased complexity of H3K36 regulation in higher eukaryotes (Wagner and Carpenter, 2012). Consistent with a more diverse role, H3K36me3 recruits a variety of effector proteins in addition to those that are recruited in yeast, including DNMT3b, which regulates gene body methylation, LEDGF, which functions in DNA repair, and ZMYND11, which regulates co-transcriptional splicing and transcription elongation (Dhayalan et al., 2010; Guo et al., 2014; Pradeepa et al., 2012; Wen et al., 2014).

The structural and functional similarities between SETD2 and Set2 provide an exceptional opportunity in which existing assays in *S. cerevisiae* can be applied to investigate the functional consequences of SETD2 mutations reported in human cancer. In this work, we characterized cancer-associated SETD2 mutations that occur at evolutionarily conserved residues in functionally important domains (i.e., the SET and SRI domains). We discovered that a missense mutation in the SET domain of SETD2 (R1625C) altered the capacity of this mutant to engage H3, leading to reduced protein stability, and a complete loss of H3K36me3. Strikingly, the same mutation in yeast Set2 (R195C) resulted in an identical effect on H3K36me3, but not H3K36me1 or H3K36me2 levels (or biological outcomes associated with these lower methylation states). Further biological studies in human cells revealed that loss of H3K36me3 in the R1625C mutant leads to DNA repair defects, thereby revealing a greater understanding of how this recurrent mutation likely leads to a loss of SETD2 tumor suppressive activity.

Results

SETD2 and Set2 Share a High Degree of Structural and Sequence Homology at their SET and SRI Domains

SETD2 and Set2 share significant structural and functional homology. SETD2 demonstrates strong sequence conservation at all of the annotated functional domains present in yeast Set2: AWS (associated with SET) 42%, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) 56%, PS (Post SET) 59%, coiled-coil 33%, WW 26%, and SRI (Set2 Rpb1 Interacting) 35% (Figure A.1A). Given this similarity, we compared the structure of the SETD2 and Set2 SET domains to identify highly conserved residues for further study. The structure of the SET domain in SETD2 has been solved by crystallography whereas the SET domain of Set2 was predicted here using I-TASSER (Roy et al., 2010; Yang et al., 2014, 2015; Zhang, 2008; Zheng et al., 2012). When the predicted structure of the Set2 SET domain was aligned with the crystal structure of the SETD2 SET domain, the structures were strikingly similar (Figure A.1B). We then examined the conservation of amino acids previously reported to be mutated in human ccRCC across six organisms (*H. sapiens*, *D. melanogaster*, *S. cerevisiae*, *M. musculus*, *X. tropicalis*, and *D. rerio*) (Cancer Genome Atlas Research Network, 2013; Dalgliesh et al., 2010; Gerlinger et al., 2012; Simon et al., 2014). Seven of the nine ccRCC mutations occur at residues that are conserved across all model organisms (Figure A.1C). Additionally, three of these seven mutations occur in a region previously identified to act as the catalytic site for lysine methylation (Sun et al., 2005). One of these mutations, R1625C, is found in a location that is adjacent to the S-adenosylmethionine (SAM) binding site in the structure, and thus would be predicted to impact catalytic activity (Figure A.1B). This residue is the most common site of missense mutation reported in both CBioPortal and COSMIC (Cerami et al., 2012; Forbes et al., 2008; Gao et al., 2013). The specific arginine to cysteine mutation is found in both glioma and ccRCC (Cancer Genome Atlas Research Network, 2013; Fontebasso et al., 2013). Significantly, mutation of the corresponding residue in *S. cerevisiae* is known to affect Set2 catalytic activity

(Jha and Strahl, 2014). Given its location and mutation frequency, we chose this mutation for further analysis.

We then examined sequence and structural conservation of the SRI domain and location of ccRCC-associated missense mutations. In contrast to the SET domain, primary sequence of the SRI domain is less conserved across model organisms (Figure A.1D). However, the aligned crystal structures of yeast and human SRI domains display structural conservation (Figure A.1E) (Li et al., 2005; Vojnic et al., 2006). In particular, the predicted site of SETD2 and RNAPII interaction was previously suggested to be the concave surface between alpha helix 1 (1) and alpha helix 2 (2) (Li et al., 2005). The physical relationship of these helices appears conserved between Set2 and SETD2. We therefore selected the R2510 residue for further study, as this amino acid is recurrently mutated (R2510H, R2510L) in ccRCC and is predicted to be important for SETD2-RNAPII interaction by *in vitro* peptide interaction assays (Cancer Genome Atlas Research Network, 2013; Li et al., 2005; Simon et al., 2014).

SET domain mutation destabilizes SETD2 in Cells

To establish a human cell system in which to study the function of SETD2 mutants, we generated SETD2 deficient cells (SETD2Δ). TAL effector nucleases (TALEN) targeting exon 3 of SETD2 were introduced into two immortalized kidney cell lines (human SV-40 immortalized proximal tubule kidney cells (HKC) and 293T) (Cermak et al., 2011; Racusen et al., 1997; Sander et al., 2011). Individual clones of TALEN-treated cells were isolated and loss of H3K36me3 was demonstrated by immunocytochemistry (Figure A.2A). We verified inactivation of both alleles of SETD2 via Sanger sequencing. Representative allelic sequencing is shown (Figure 2B).

We then exogenously expressed a truncated wild-type FLAG-tagged form of SETD2 (amino acids 1323-2564; tSETD2), which includes all known functional domains. The R1625C or R2510H mutants were generated in tSETD2. Relative to tSETD2 and R2510H, R1625C

protein levels were reduced (Figure A.2C). R1625C mutant mRNA levels were also less abundant (Figure A.2D). We examined protein stability after treatment with the protein synthesis inhibitor cycloheximide. The R1625C protein demonstrated a significantly shorter half-life compared to that of wild-type (Figure A.2E). In contrast, the half-life of the R2510H mutant was unchanged (Figure A.2E). These data suggest that the decreased protein level of the R1625C SET domain mutant results from both decreased RNA and a shortened protein half-life.

Histone H3 Lysine 36 Trimethylation is Linked to SETD2 Mutational Status

We interrogated H3K36 methylation status in cells transiently transfected with either tSETD2 or the mutants, R1625C and R2510H. Using immunocytochemistry (ICC) we found that transfection of tSETD2 resulted in global restoration of H3K36me3 (Figure A.2F), demonstrating that the N-terminus is not required for catalytic activity of SETD2. Transfection of R1625C (SET domain) mutant construct failed to restore H3K36me3. In contrast, expression of the R2510H SRI mutant globally restored H3K36me3.

We next examined the H3K36 methylation status by western blot analysis. Consistent with findings from ICC, SETD2Δ cells show complete loss of H3K36me3. Trimethylation was restored to wild-type levels by expression of either the tSETD2 or the SRI mutant. In contrast, the SET domain mutant failed to trimethylate H3K36 (Figure A.2G). Monomethylation (H3K36me1) and dimethylation (H3K36me2) were unaffected by SETD2 loss or expression of SETD2 variants (Figure A.2G). These results are in agreement with the findings that SETD2 is the exclusive H3K36 trimethyltransferase in mammalian cells.

We then asked whether expression of either tSETD2 or R2510H restored H3K36me3 to levels similar to wild-type cells at specific loci. H3K36me3 levels have been shown to increase along the gene body with preference for exons (Kolasinska-Zwierz et al., 2009). Using ChIP-qPCR, we examined the H3K36me3 levels at multiple exons of two genes, CDK2 and MYC, which had previously been described (Carvalho et al., 2013). As expected, SETD2Δ cells displayed low H3K36me3 levels at all sites. (Figure A.2H). Expression of tSETD2 recapitulated

the previously described pattern for H3K36me3 in wild-type cells at both CDK2 and MYC, with higher signal at exons 5 and 6 relative to exon 1 in CDK2, and in exons 2 and 3 relative to exon 1 in MYC (Carvalho et al., 2013). Cells expressing the R1625C SET domain mutant displayed loss of H3K36me3 at levels similar to that of SETD2Δ cells. Finally, expression of the R2510H mutant also showed greater signal at later exons, indicating that this point mutation restores not only the levels of methylation, but the spatial placement of these methyl marks on actively transcribed genes.

The SETD2 R1625C Variant is Enzymatically Inactive *in vitro* and has Diminished Substrate Binding

Given the R1625C SETD2 variant is associated with loss of H3K36me3 in cells, we asked whether the R1625C mutation disrupts the methyltransferase activity of SETD2 *in vitro*. To do this, we expressed and purified from bacteria a wild-type or R1625C mutated fragment of SETD2 (residues 1345-1711) containing the SET domain. Both the wild-type and the R1625C SET domain constructs yielded soluble proteins that were >90% pure as assessed by SDS-PAGE (Figure A.3A). Methyltransferase activity was then assessed using a radiometric assay with chicken oligo-nucleosomes as the substrate. Whereas wild-type SETD2 displayed robust activity, the R1625C variant displayed little enzymatic activity over the no enzyme control (Figure A.3B).

We next sought to determine why the R1625C variant is catalytically inactive. We first considered whether this mutation results in a misfolded protein, thereby inactivating the SET domain. We compared the circular dichroism (CD) spectra of the wild-type SETD2 with the R1625C variant. The CD spectra in the low UV range (185-260nm) of the wild-type and the R1625C variant were nearly indistinguishable, suggesting that the R1625C substitution does not alter the secondary structure of the SET domain (Figure A.3C). To determine if the R1625C variant alters the thermal stability of the SET domain, we monitored the CD signal at the 207 nm

peak over a temperature range from 20-95°C. Both the wild-type and the R1625C variant showed highly similar thermal melt curves with a melting temperature (Tm) of approximately 55°C (Figure A.3D). Together, these results suggest that the loss of catalytic activity observed for the R1625C variant is not due to protein misfolding or reduced thermal stability.

Structural analysis of the SETD2 SET domain shows that R1625 is positioned within the active site, opposite the SAM binding pocket, and is located about 7Å away from the sulfur group of S-adenosyl-homocysteine (SAH) (Figure A.3E). While substitution of R1625 with cysteine would not be expected to directly disrupt SAM binding, the R1625 side chain engages in three hydrogen bonds with the backbone carbonyl oxygens of A1617 and T1618 (Figure A.3E). Substituting cysteine for R1625 using in silico mutagenesis showed that every possibly cysteine rotamer would cause steric clashes. The cysteine side chain would not recapitulate the hydrogen bonding network of R1625 when oriented in the same direction as the R1625 side chain observed in the crystal structure (Figure A.3F). Although no structure of the SETD2 SET domain ternary complex containing histone H3 is available, the location of R1625 in close proximity to, but opposite the SAM binding pocket suggests that R1625 may directly or indirectly engage H3 or may maintain local structural integrity that aids substrate binding.

To determine if the R1625C variant has altered substrate binding, we performed peptide pull-down experiments using histone H3 peptides that were unmodified, or methylated at K36. The pull-down experiments showed that the R1625C variant associated with all of the histone peptides to a lesser degree compared to the wild-type SETD2 SET domain, suggesting that the R1625C substitution weakens substrate binding (Figure A.3G). Taken together, our results suggest that the R1625C substitution impairs enzymatic activity by reducing substrate binding, which is likely a consequence of fine structural disturbances induced by loss of the R1625 hydrogen bonding network within the active site.

Domain-Specific Mutations in Yeast Set2 Separate Roles of H3K36 Methylation States

To further explore the functional significance of ccRCC-associated mutations, we took advantage of several well-characterized phenotypic assays in *S. cerevisiae*. Using *set2Δ* cells, which are devoid of all H3K36 methylation (Jha and Strahl, 2014), we created strains that either contained vector alone, or strains that exogenously express either wild-type or mutated forms of Set2. Mutant forms of Set2 included the homologous SETD2 SET domain mutant (R195C), the homologous SETD2 SRI mutant (K663L), or a control SET domain mutant (H199L) previously characterized to disrupt both tri- and di- methylation, while retaining monomethylation activity (Jha and Strahl, 2014). As expected, Set2 loss resulted in the complete absence of mono-, di-, and trimethylation of H3K36, which was rescued upon addition of wild-type *SET2* (Figure A.4A). As previously shown, the H199L mutant only restored monomethylation. In contrast, while the K663L mutant restored all H3K36 methylation states, the R195C only restored H3K36 mono- and dimethylation. Intriguingly, the restoration of H3K36 mono- and dimethylation by the R195C mutant mimics the status of SETD2 deficient human cells (i.e., both have a selective loss of H3K36me3). Since the SETD2 R1625C mutant demonstrated decreased protein stability in human cells, we examined protein levels of the R195C mutant in the yeast cells. Following cycloheximide treatment we observed decreased protein levels of the R195C mutant relative to wild-type Set2, particularly at 3 hours post-treatment. This effect was rescued by treatment with the proteosome inhibitor MG132 (Figure A.4B). This suggests that, like in humans, the R195C variant is less stable than wildtype in yeast cells.

Loss of Set2 has been implicated in various phenotypes in *S. cerevisiae*, including transcription elongation defects, cryptic initiation and sensitivity to DNA damaging agents (Jha and Strahl, 2014). We asked whether the R195C Set2 mutant would be associated with any of these phenotypes. To examine transcriptional elongation, we performed a spotting assay in the presence of the transcription elongation inhibitor 6-Azauracil (6-AU). This assay has been previously used to assay for the presence of transcriptional elongation defects in yeast (Kizer et

al., 2005). As expected, wild-type yeast were sensitive to 6-AU (200 µg/mL), whereas *set2Δ* cells were resistant to this drug (Figure A.4C) (Kizer et al., 2005). While expression of wild-type *SET2* restored sensitivity to 6-AU, the H199L mutant did not. The R195C and K663L mutants behaved similar to wild-type Set2. These data suggest that H3K36me2 is primarily responsible for the sensitivity to inhibitors of transcriptional elongation.

Cryptic initiation has been previously associated with Set2 loss (Lickwar et al., 2009). We therefore assessed the effects of our Set2 mutations in a cryptic transcription reporter assay. This assay monitors the growth of yeast cells that contain the *HIS3* gene with a cryptic start-site that exists in the *FLO8* gene. Importantly, the cryptic start site is out of frame when the 5' promoter is used and a functional transcript is only produced if the 3' cryptic start site is utilized. In this setting, cryptic transcription results in expression of *HIS3,* which can restore growth in medium lacking histidine. Consistent with previous results, loss of Set2 permits growth in the absence of histidine (Figure A.4D) (Lickwar et al., 2009). No growth was observed in the cells expressing the R195C or K663L mutants. However, cell growth occurred in the presence of the H199L mutant (Figure A.4D). These data indicate that trimethylation is dispensable for preventing cryptic initiation, whereas dimethylation is required to suppress this phenotype.

Recent studies have demonstrated that yeast lacking Set2 cannot properly activate the DNA-damage checkpoint (Jha and Strahl, 2014; Pai et al., 2014). To determine if the RCC-associated SET domain mutation impacts this phenotype, we assessed the impact of the Set2 point mutants on growth in the presence of phleomycin, a double-strand break-inducing agent. As expected, *set2Δ* cells displayed increased sensitivity to phleomycin relative to Set2 wild-type yeast (Figure A.4E). *set2Δ* cells expressing either wild-type Set2 or the R195C or K663L mutants showed similar sensitivity as the wild-type rescue. However, yeast expressing the H199L mutant showed a similar level of sensitivity as the *set2Δ* cells (Figure A.4E). Taken together, these data indicate that the cellular phenotypes associated with Set2 loss in yeast are

associated with H3K36me2, and that H3K36me3 is dispensable for these activities (summarized in Figure A.4F).

Human Kidney Cells Display an H3K36 Trimethylation-Dependent DNA Damage Response

Because of the exclusivity of SETD2 in mediating trimethlyation in human cells, we studied similar phenotypes to those examined in yeast in human cells that express ccRCC-relevant mutants. We first examined the effect of the transcriptional elongation inhibitor 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) on cell survival. DRB inhibits CDK9, which results in premature termination of transcription (Zhu-Yr et al., 1997). Assessing viability at 12 hour time points for 3 days, we observed that DRB-associated toxicity did not differ between SETD2 wild-type and SETD2Δ cells (Figure A.5A).

Several recent studies have examined the effect of SETD2 loss in human cells on the response to DNA damage (Aymard et al., 2014; Carvalho et al., 2014; Daugaard et al., 2012; Kanu et al., 2015; Pfister et al., 2014). To further explore the role of H3K36me3 in the DNA damage response, we irradiated HKC cells to 2 gray and then performed immunofluorescence for γH2A.X, a marker of DNA damage. At 30 minutes post irradiation, γH2A.X foci were seen in all cell types at similar levels (Figure A.5B). In untransfected and in control transfected wild-type cells, the number of foci greatly decreased by 1 hour and largely resolved by 4 hours. However, in SETD2Δ cells, the number of γH2A.X foci remained elevated at both 1 hour and 4 hours. Expression of tSETD2 in SETD2Δ cells led to resolution of foci at time points similar to wild-type cells. Cells expressing the SRI mutant, R2510H, also showed rapid foci resolution. However, foci persisted in cells expressing the R1625C mutant (Figure A.5B). Quantification of these results demonstrated that both SETD2Δ cells and R1625C expressing cells had a significantly higher percentage of cells with greater than 10 foci compared with the other conditions (Figure A.5C).

We quantified γH2A.X by immunoblotting, enabling us to account for changes in total protein and histone levels. These studies were performed in 293T cells, as additional validation of results in HKC cells. As observed with HKC cells, regardless of SETD2 status, 293T cells showed increased total γH2A.X at 30 minutes post irradiation (Figure A.5D). By 4 hours γH2A.X levels returned to baseline in cells with H3K36 trimethylation (WT, tSETD2, R2510H). However, elevated levels of γH2A.X were observed in cells lacking H3K36me3 associated with SETD2 loss or R1625C expression. Finally, we examined the effect of irradiation on viability using a colony formation assay. The fraction of surviving colonies did not differ between SETD2 wild-type and SETD2Δ cells (Figure A.5E). Overall, these findings demonstrate that SETD2-mediated H3K36me3 is coupled to the efficient resolution of double-strand breaks. Corresponding to results in yeast, loss of trimethylation is not associated with enhanced sensitivity due to inhibition of transcriptional elongation or from DNA damage.

Discussion

In an effort to explore the function of missense mutations identified in human cancers, we examined several recurrent mutations that occur at evolutionarily conserved residues in yeast and human cell lines. The striking homology between SETD2 and Set2 creates an opportunity to compare the effects of mutations while taking advantage of the strengths of each model system. Indeed, a recent study also modeled cancer mutations in yeast, highlighting the utility and power of yeast to be a robust model system to aid in human protein analyses (Sun et al., 2016). In this paper, we investigated how two highly conserved SETD2 residues that are commonly mutated in cancer affect the functions of this enzyme. Limited studies have explored the potential roles that SETD2 loss may play in cancer development. We found that mutation of the SET domain, but not the SRI domain, resulted in effects in the human and yeast assays. Specifically, we identified R1625C as a critical mutation that impacted SETD2 enzymatic activity and protein stability in cells –an effect also noted when this mutation was modeled in yeast

Set2. Loss of H3K36me3 in human cells also led to defects in DNA repair, indicating a potential mechanism by which SETD2 functions as a tumor suppressor. To our knowledge, this study is the first to dissect the impact of cancer-associated mutations in SETD2, and further validate using yeast as a model to complement human cell analyses.

A key discovery emerged from the study of the R1625C mutant. In contrast to the human R1625C variant which was catalytically inactive *in vitro*, the homologous substitution in yeast Set2 led to an uncoupling of di- and trimethylating activities. This suggests that this residue may be important for the specific trimethylating activity of the enzyme. Moreover, many substrate binding interactions govern stability. Thus, the reduced protein stability (in the absence of other thermal instability) may reflect a structural role that differentiates mono-, di-, and trimethylating activity. Because of this unique mode of regulation, the R195C mutation allowed us to examine the functions specifically associated with the trimethylated state of H3K36 in cells (i.e., impaired transcriptional elongation, cryptic initiation, and impaired survival in the face of DNA damage). Consistent with other reports that examined cryptic initiation, we found that H3K36me3 is dispensable whereas H3K36me1/me2 is required to suppress cryptic initiation, as well as for transcription elongation and DNA damage survival phenotypes (Li et al., 2009; Youdell et al., 2008). In contrast, the SET domain mutation in SETD2 had a similar impact on H3K36me3 levels but resulted in a clear DNA damage response phenotype.

These studies offer a rationale for differences in observed phenotypes in SETD2 deficient human cells associated with Set2 loss in yeast, including cryptic initiation and impaired transcriptional elongation. However, impaired response to DNA damage, as we observed, has been reported for both mammalian systems and Set2 in yeast, linking this feature with H3K36 trimethylation. The yH2A.X results suggest that resolution of DNA strand breaks in human cells requires H3K36me3. Due to the presence of multiple H3K36 dimethylating enzymes in mammalian cells, absence of H3K36me2 is rarely encountered in human models.). H3K36me2 is induced by ionizing radiation and improves association of early DNA repair components with

an induced break, and improved repair by non-homologous end joining (NHEJ) in human cells (Fnu et al., 2011). Although our data show that the resolution of strand breaks, as measured by clearance of γH2A.X foci, was delayed in the absence of H3K36me3, our data also shows that H3K36me3 loss does not affect viability after radiation in mammalian cells. Thus, loss of dimethylation may convey a sensitivity to DNA damage that is not present in the absence of SETD2 trimethylating activity. It is important to consider that multiple factors may influence cell death in transformed cells. However, these distinct findings in yeast and mammalian systems indicate a complex level of regulation of DNA repair mediated by the histone code at H3 lysine 36. Multiple studies have concluded that the loss of SETD2 confers a variety of types of genomic instability, ranging from microsatellite instability to impairment of NHEJ (Awwad and Ayoub, 2015; Li et al., 2013; Pai et al., 2014). Our data agree with these results.

Through modeling disease-relevant SETD2 mutations, we were able to gain insight into H3K36me3 function and dissect the roles of H3K36 dimethylation and trimethylation. Future studies will further explore the roles of the SETD2 SRI domain and examine the effects of additional mutations, and will further define the role of SETD2 loss in the development of kidney cancer and other tumor types.

Materials and Methods

Modeling SETD2 and Set2

The primary protein sequences of Set2 from *Saccharomyces cerevisiae* and SETD2 from *Homo sapiens* were compared via BLAST alignment analysis and the percentage of homology between annotated domains was determined using the percentage overlap of the BLAST-aligned regions. The primary sequences of the SET and SRI domains of the enzyme responsible for H3K36 methylation from *H. sapiens*, *S. cerevisiae*, *X. tropicalis*, *D. melanogaster*, *D. rerio*, and *M. musculus* were aligned using ClustalOmega, and annotated with reported SETD2 mutations in ccRCC (Cancer Genome Atlas Research Network, 2013;

Dalgliesh et al., 2010; Gerlinger et al., 2012; Sievers et al., 2011; Simon et al., 2014). The structure of the SETD2 SET domain, SETD2 SRI domain and Set2 SRI domain haves been previously reported reported (Li et al., 2005; Vojnic et al., 2006; Zheng et al., 2012). To predict the structure of the yeast SET domain, the amino acid sequence (UniProtKB, P46995) was submitted to I-TASSER using the default parameters (Roy et al., 2010; Yang et al., 2014, 2015; Zhang, 2008). The ribbon structures were aligned using the align command in the PyMOL Molecular Graphics System (Schrödinger, LLC, 2015).

Mammalian Cell Lines Transfections and Phenotypic Assays

293T human embryonic kidney cells were generously provided by Dr. Jenny Ting, Chapel Hill, NC. The SV-40 transformed human renal tubule epithelial cell line (referred to as HKC) was obtained from Dr. Lorraine Racusen, Baltimore, MD (Racusen et al., 1997). A pair of vectors containing TAL effector nucleases (TALENs) targeting exon 3 of SETD2 was generated using the REAL (Restriction Enzyme And Ligation) assembly method. Component plasmids were obtained from Addgene (www.addgene.org/talengineering/talenkit/). Briefly, target sites were selected and TALENs designed using Zifit (http://zifit.partners.org/ZiFiT/), followed by assembly (Sander et al., 2011). The TALEN target sequences are: 5'-TCATGTAACATCCAGGCC -3' and 5'-ACAGCAGTAGCATCTCCA-3'.

An expression construct containing a N-terminal truncated form of SETD2 (amino acids 1323 to 2564; tSETD2) was sequence-optimized for expression in human cells, tagged with the FLAG sequence on the C-terminus, and synthesized by Life Technologies. tSETD2 was specifically used as it modelss the yeast protein in domain structure, and expression of full-length SETD2 was technically unfeasible. tSETD2 was subcloned into the pINDUCER20 vector(58). Disease-relevant SETD2 mutations were introduced into the tSETD2 pINDUCER20 construct using the QuikChange II Site-Directed Mutagenesis Kit according to the manufacturer's instructions (Agilent Technologies). Mutations were verified through direct DNA sequence analysis.

293T and HKC human renal cells were transfected with the TALEN constructs, tSETD2 construct, and mutation constructs using Amaxa® cell Line Nucleofector® Kit V (Lonza). For the protein stability assay, cycloheximide (100 ng/mL) was applied to cells 72 hours post-transfection. For the 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) transcription inhibition assay, 1000 cells/well were plated in triplicate on a 96 well plate and treated with 100uM DRB for 72 hours, with viability being measured every 12 hours by Cell Titer Glo (Promega).

Sequencing and Allelic Analysis

DNA was extracted and the SETD2 TALEN target site was PCR-amplified (primers: 5'-ACAGGGACGACAGAAGGTGTCATT-3' and 5'-ACTGGTGCTGGTGATGAGAGTGTT-3'), sequenced (Applied Biosystems 3730xl Genetic Analyzers, Life Technologies) and analyzed (Sequencher DNA analysis software version 5.0, Gene Codes Corporation). Allelic analysis was performed by subcloning individual PCR products (TOPO TA Cloning® Kit, 45-0641, Life Technologies). DNA from individual clones was PCR amplified, sequenced and analyzed as described above.

Immunoblot Analysis

To isolate mammalian cellular proteins, cells were lysed in Mammalian Protein Extraction Reagent (M-PER; Pierce Biotechnology) supplemented with Complete Mini Protease Inhibitor Cocktail (Roche). Histones were extracted using an overnight acid extraction protocol (Abcam). For yeast immunoblots, asynchronously grown mid-log (0.6-0.8 OD) phase cultures were lysed by SUMEB using glass beads methods described on the Gottschling Lab website (http://labs.fhcrc.org/gottschling/Yeast%20Protocols/pprep.html).

Antibodies

Antibodies used include: SETD2 (HPA042451, Sigma-Aldrich, St. Louis, MO), Ku80 (ab3107, Abcam), H3K36me3 (ab9050, Abcam), total H3 (Abcam, ab10799; Epicypher, 13-0001), H3K36me2 (39255, Active Motif), H3K36me1 (ab9048, Abcam), γH2AX (ab2893; Abcam), GST

(EpiCypher # 13-0022), and Set2 (raised in Strahl lab). Secondary antibodies used in human studies were anti-mouse and anti-rabbit IRDye Secondary Antibodies from LI-COR Biosciences (Lincoln, NE). HRP-conjugated donkey anti-Rabbit secondary antibody was used (Amersham) for yeast studies. Human antibodies were detected using the Odyssey IR imager (LICOR Biosciences) and densitometry analysis was performed using ImageStudio Ver2.0. The yeast immunoblots were developed using ECL-Prime (Amersham) and densitrometry analysis was done using ImageJ (NIH).

Immunocytochemistry

Cells were fixed with 4% para-formaldehyde for 15 minutes and permeabilized using 0.25% Triton X-100 in PBS. Endogenous peroxidase activity was blocked by incubation in 1% H_2O_2. Cells were then blocked in 5% bovine serum albumin followed by incubation in primary antibody. The Vectastain ABC Kit (PK6101, Vector Laboratories) was used for secondary antibody and HRP conjugation followed by the DAB peroxidase substrate kit (SK-3100, Vector Laboratories) and hematoxylin staining.

Chromatin Immunoprecipitation

Cells were fixed in 1% formaldehyde for 10 minutes, quenched with 125 mM glycine treatment, and homogenized in hypotonic solution (10 mM Tris pH 7.4; 15mM NaCl; 60mM KCl; 1mM EDTA; 0.1% NP-40; 5% sucrose; 1x protease inhibitors). Nuclei were separated by centrifugation through a sucrose pad (10mM Tris pH 7.4; 15mM NaCl; 60mM KCl; 10% sucrose; 1x protease inhibitors) then resuspended in ChIPs buffer (10mM Tris pH 7.4; 100mM NaCl; 60mM KCl; 1mM EDTA; 0.1% NP-40; 1x protease inhibitors, 0.05% SDS) and sonicated to obtain DNA between 200 bp to 1 kb. DNA was immunoprecipitated with H3K36me3 antibody prebound to protein A/G beads. Immunoprecipitated complexes were washed, RNAse and Proteinase K treated, and protein-DNA cross-links were reversed by overnight incubation at 65°C.

Quantitative RT-PCR

Total RNA was extracted using the Qiagen RNeasy mini kit. cDNA was made from total RNA using random primers and Superscript II Reverse Transcriptase reagents (Invitrogen).

Expression and Purification of Human SETD2

An *E. coli* codon-optimized synthetic gene corresponding to human SETD2 (UniProtKB ID Q9BYW2) residues 1345-1711 followed by a stop codon was cloned into the pGEX-6P-2 expression vector (GE Healthcare) using standard procedures. The protein was expressed in soluBL21 (DE3) (Amsbio) cells by growing cells in Terrific Broth II media (MP Biomedicals) at 37°C until an $0D_{600}$ of ~0.6 then chilling the cells for 30 minutes at 4°C before inducing them with 1mM IPTG in the presence of 25 μM ZnCl2 for 20 hours at 16°C. Cells were harvested by centrifugation and pellets were flash frozen in liquid nitrogen. For purification, thawed cell pellets were resuspended in binding buffer (50mM Tris pH 7.3, 300mM NaCl, 4mM dithiothreitol (DTT), 10% glycerol and 1 μM ZnCl2) supplemented with 1 Complete mini-EDTA-Free Protease Inhibitor Tablet (Roche), 0.1mM phenylmethane sulfonyl fluoride (PMSF), 0.5mg/mL chicken egg lysozyme, and 0.2 % (v/v) Triton X-100 and incubated on ice for 45 minutes, then lysed with sonication and clarified by centrifugation. Clarified lysates were diluted 1:2 with binding buffer and applied to a 5mL glutathione agarose gravity flow column (pre-equilibrated with 10 column volumes (CV) of binding buffer) at a flow rate ~ 0.5mL/min at 4°C. The bound protein was washed with 10CV of binding buffer then eluted from the column with 35 mLs of elution buffer (50mM Tris pH 8.0, 300mM NaCl, 4mM DTT, 10% glycerol, and 10mM reduced L-glutathione). The eluted protein was mixed with Precision Protease (GE Healthcare) and exhaustively dialyzed against binding buffer, without ZnCl2, over the course of 20 hours at 4°C. The cleaved protein sample was applied to a pre-equilibrated 5mL glutathione agarose gravity flow column at a flow rate ~ 0.5mL/min at 4°C and the flow-through was collected and concentrated using an Amicon-Ultra 15 concentrator (Millipore). The Bradford Assay and and SDS-PAGE analysis were used to determine the quantity and purity of the protein samples respectively. The SETD2

R1625C mutant was generated by site-directed mutagenesis using the QuickChange Kit (Agilent), and expressed and purified as described above. Note: a small amount of GST-SETD2 WT and R1625C was not treated with Precision Protease, but was extensively dialyzed against binding buffer then used for peptide pull-down experiments (see below).

Histone Methyltransferase Assays

HMT assays were preformed by incubating wild type SETD2 or the R1625C variant at a final concentration of 500nM with 1 µg of chicken oligo-nucleosomes (EpiCypher), and 1 µCi 3H-AdoMet (PerkinElmer Life Sciences) in a buffer containing 50mM HEPES pH 8.0, 150mM NaCl, 2.5mM $MgCl_2$, 1µM $ZnCl_2$, and 2.5% glycerol, for 16 hours at room temperature (total reaction volume was 20µL). The reactions were quenched with 0.5% TFA then spotted onto Whatman filter paper, air dried and washed 4 times with ~200 mLs of a sodium bicarbonate (pH 9.0) solution, air dried again and added to liquid scintillation vials containing 5 mLs of Ultima Gold F (PerkinElmer Life Sciences). Samples were counted for 1 minute each using an all-purpose Beckman Coulter liquid scintillation counter in 3H mode. A reaction without enzyme was used as the negative control and to determine background counts.

Circular Diochroism Spectroscopy

For CD experiments, proteins were exhaustively dialyzed into a buffer containing 20mM sodium phosphate (pH 7.0), 150mM sodium fluoride (NaF), and 0.2mM tris(2-carboxyethyl)phosphine (TCEP) at 4°C. CD spectra were collected using a 0.1cm quartz cuvette and a Chirascan Plus instrument (Applied Photophysics Inc.) at 20 °C over the wavelength range 185-260nm with a step size of 0.5nm. A sample of the buffer was collected over the same wavelength scan and absorbance values were subtracted from the final datasets. Each sample was scanned three times and the final plots represent the average scan with the CD signal (in milidegrees) converted to molar ellipticity ([Θ]). For thermal melt curves, the CD absorbance at 207nm was collected over the temperature range from 20-95°C with 1°C temperature ramping and a

temperature tolerance range set to 0.2°C. Proteins were diluted to 0.25mg/mL for all CD data collection (protein stock concentrations determined by A_{280}).

Peptide Pull-Downs

A total of 50pmols of GST tagged-wild type SETD2 or the R1625C variant was incubated with 500pmols of each biotinylated histone peptide for 1 hour at 4°C in peptide binding buffer (50mM Tris pH 8.0, 300mM NaCl, 0.1% NP-40) supplemented with 2mM Dithiothreitol (DTT) and 1 μM ZnCl2. Following incubation, the protein-peptide mixture was incubated with streptavidin coated magnetic beads (Pierce), pre-equilibrated with peptide binding buffer, for 1 additional hour at 4°C. The beads were washed 3 times with peptide binding buffer and bound complexes were eluted with 1x SDS loading buffer, resolved via SDS-PAGE and transferred to a PVDF membrane. The membrane was probed with anti-GST antibody diluted to 1:4000 in 5% BSA in PBS-T. The peptides contained the budding yeast histone H3 residues 27-45 and were mono-, di-, or trimethylated at Lys36. In this region, the human and budding yeast H3 sequences differ by an Ala to Ser substitution at residue 31 and by an Arg to Lys substitution at residue 42.

Yeast Growth Assays

Parental yeast strains were transformed with indicated plasmids, and were grown to saturation in appropriate selection medium. Saturated cultures were diluted to an OD_{600} of 0.5 and 5-fold serially diluted and plated with or without 6-azauracil (6-AU) or plated with or without phleomycin; pictures were taken 2-3 days after spotting. Similarly, strains with integrated cryptic initiation cassette (as shown in Figure A.3) were serially diluted and plated on SC-Ura-His plates with or without galactose for 3 days to detect growth. Growth on SC–Ura acted as a control for equivalent growth for all the strains.

Immunofluorescence Staining for γH2A.X

HKC cells were cultured for 16-18 hours followed by 2 Gy radiation (RS 2000 Biological Research Irradiator), fixed for 15 minutes in 4% paraformaldehyde, washed with cold PBS, and permeabilized (0.25% Triton X-100 in PBS) for 10 minutes. After blocking with 1% hydrogen

peroxide then 5% BSA, cells were incubated with anti-γH2AX (1:500) at 4°C overnight. Cells were washed using PBST and incubated with goat anti-rabbit IgG (1:500) Cy5 1h at RT. After washing, cells were counterstained using DAPI. Fluorescence signals were visualized using confocal microscopy (LSM 700, Zeiss) and number of foci per cell were analyzed using Zen (LSM 700, Zeiss). 5 images per coverslip (total 15 images) were collected in three independent experiments. For the radiation colony formation assay, cells were diluted to a single cell suspension and 300 cells were plated on a 10 cm plate. Plates were irradiated at 0, 37, 75, 150 and 300 rads, allowed to grow for 10 days then stained with crystal violet. Colonies were counted manually.

Figures

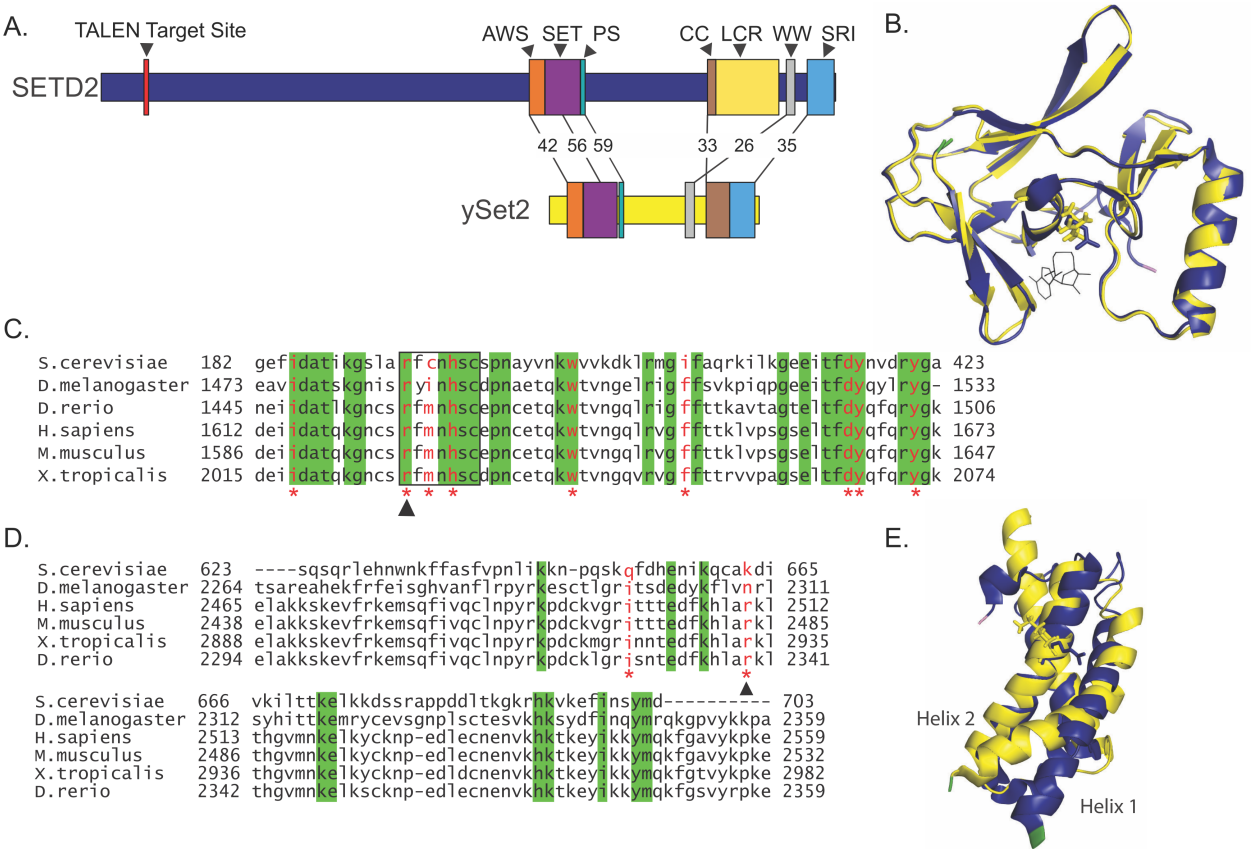


Figure A.1: SETD2 and yeast Set2 show high sequence and structural conservation

A. Comparison of SETD2 and yeast Set2 (ySet2) annotated protein structure. Percentage of conserved residues within the BLAST aligned domain sequence is indicated. Annotated domains include: AWS: Associated with SET, SET: Su(var)3-9, Enhancer-of-zeste, Trithorax, PS: Post-SET, CC: Coiled-Coil, LCR: Low Charge Region, WW: conserved Trp residues, SRI: Set2 Rpb1 interacting. Numbers represent percent conservation. B. Alignment of human SET domain crystal structure (blue) with I-TASSER protein structure prediction for yeast SET domain (yellow). N-terminus is marked in green, C-terminus is marked in pink, and residues mutated are shown as sticks. C. Partial SET domain sequence alignment across multiple species. Amino acids 1612-1673 of human SET domain (amino acids 1550-1667) are shown. Residues mutated in ccRCC are in red and marked with an asterisk. The arrow indicates R1625, the residue mutated for study. The black box indicates residues previously shown to be an important catalytic site. Residues that are conserved across species are indicated in green. D. SRI domain sequence alignment across multiple species. Residues mutated in ccRCC are in red and marked with an asterisk. The arrow indicates R2510, the residue mutated for study. Residues that are conserved across species are indicated in green. E. Alignment of human SRI domain crystal structure (blue) with yeast SRI domain crystal structure (yellow). N-terminus is marked in green, C-terminus is marked in pink, and residues mutated are shown as sticks.

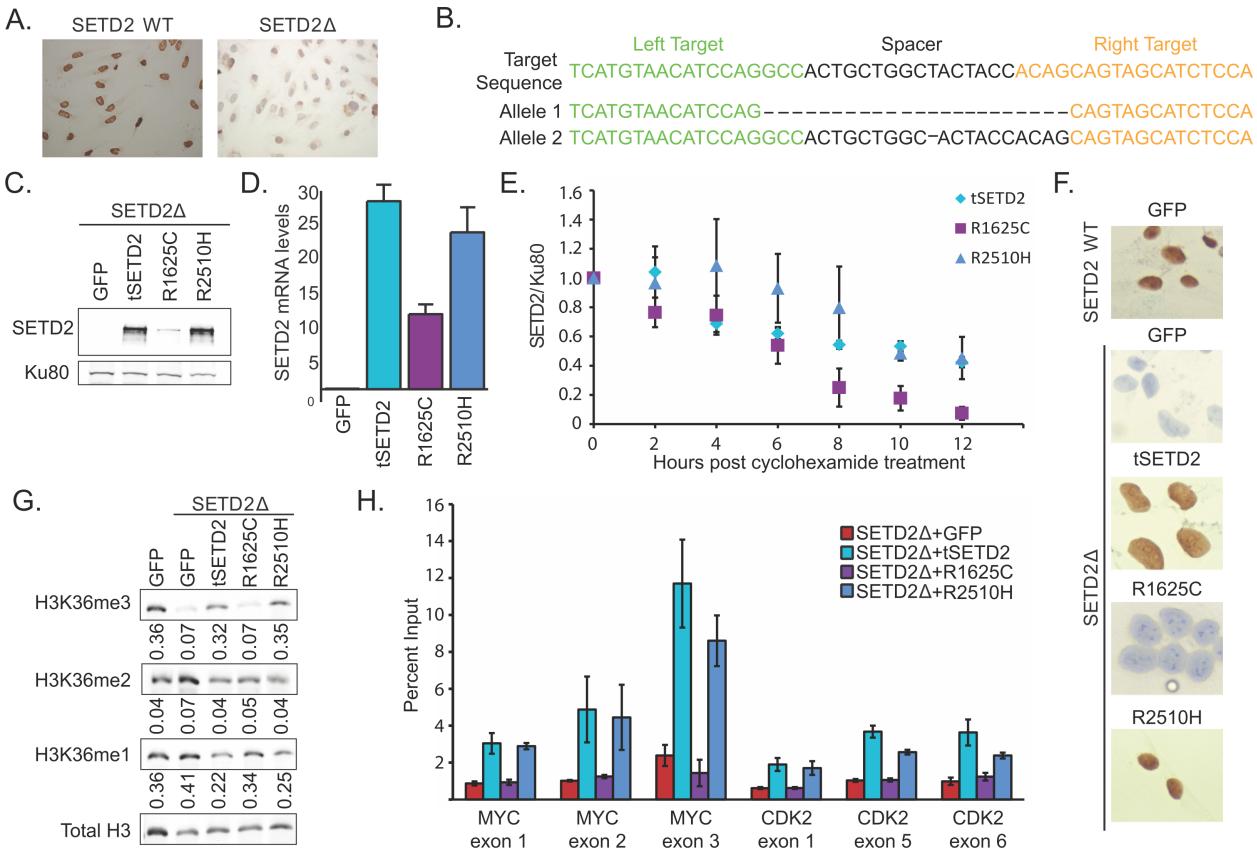


Figure A.2: ccRCC specific mutations in SETD2 have separate effects on H3K36me3

A. Immunocytochemistry of HKC SETD2 wild-type (top) and SETD2Δ cells for H3K36me3. B. Sanger sequencing results of TALEN target sequence in exon 3 of SETD2. Two allelic variants in one HKC SETD2Δ clone are represented. C. Immunoblot displaying protein expression level 72 hours after transfection in 293T cells. Ku80 acts as a loading control. D. Average quantification SETD2/Ku80 over the hours of 12 hours following cycloheximide treatment in three independent western blots (left). Average half-life of mutant SETD2 proteins (right). E. Average RNA levels of tSETD2, R1625C, or R2510H, as determined by qPCR for tSETD2 levels. F. Anti-H3K36me3 immunocytochemistry on HKC cells at 72 hours post-transfection following reintroduction of GFP, wild-type tSETD2, R1625C, or R2510H. G. Anti-H3K36 methylation immunoblot displaying levels of methylation levels at 72 hours post-transfection following reintroduction of GFP, wild-type tSETD2, R1625C or R2510H. Quantification of H3K36me3/H3 levels is shown beneath blot. H. ChIP-qPCR displaying H3K36me3 levels at exonic locations in CDK2 (left) and Myc (right), displayed as ChIP signal/Input. Error bars represent standard error. Significance comparisons were made to SETD2 inactive + GFP.

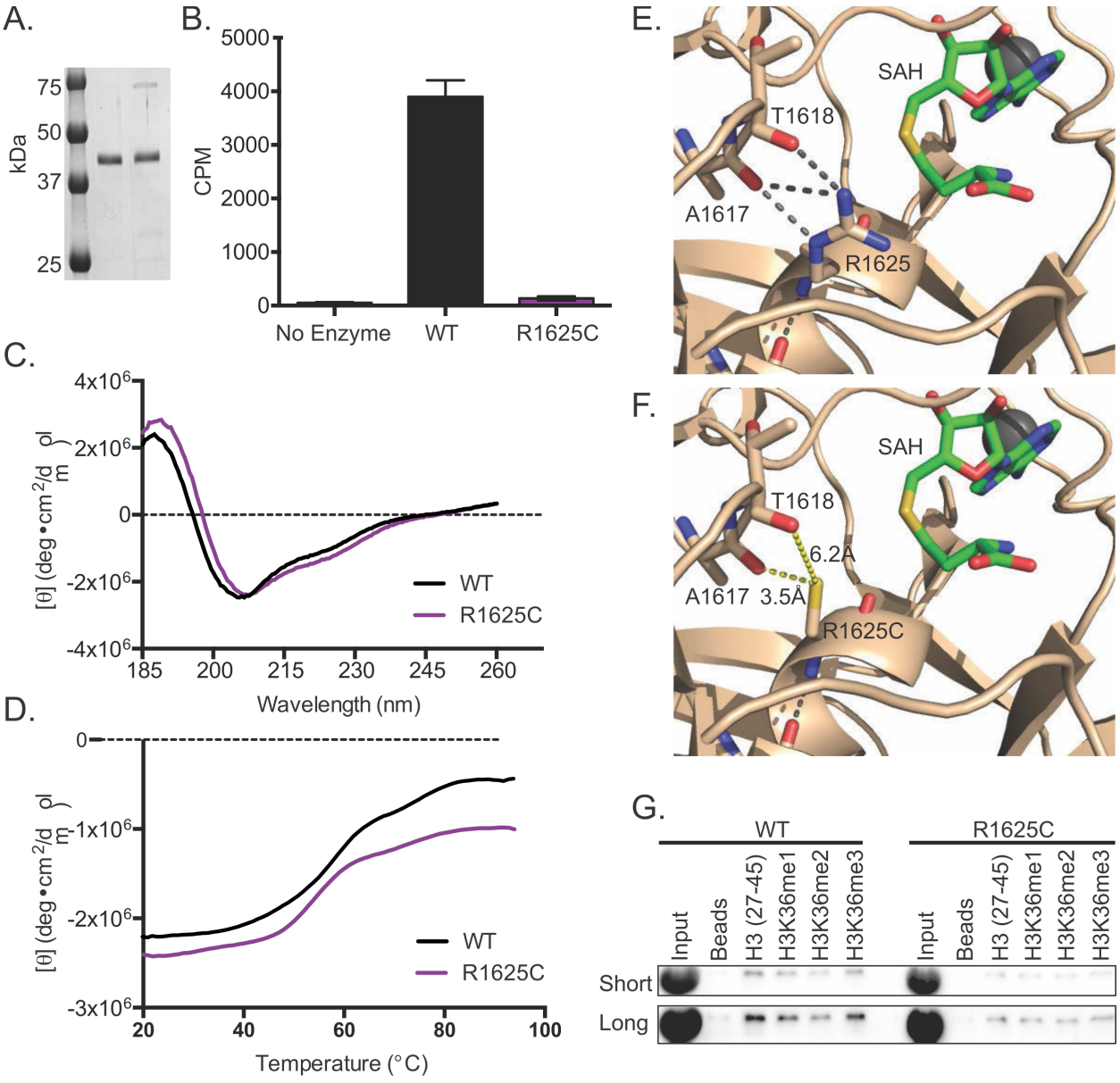

A.
kDa
75
50
37
25
B.
CPM
5000
4000
3000
2000
1000
0
No Enzyme
WT
R1625C
C.
[θ] (deg•cm²/dmol)
4x10⁶
2x10⁶
0
-2x10⁶
-4x10⁶
185
200
215
230
245
260
Wavelength (nm)
WT
R1625C
D.
[θ] (deg•cm²/dmol)
0
-1x10⁶
-2x10⁶
-3x10⁶
20
40
60
80
100
Temperature (°C)
WT
R1625C
E.
SAH
T1618
A1617
R1625
F.
SAH
T1618
6.2Å
A1617 3.5Å
R1625C
G.
WT
R1625C
Input
Beads
H3 (27-45)
H3K36me1
H3K36me2
H3K36me3
Short
Long

Figure A.3: The SETD2 R1625C variant is catalytically inactive and has reduced substrate-binding capacity

A. Coomassie Blue stained SDS-PAGE gel of 1µg of purified wild type or R1625C variant SETD2 construct containing amino acids 1345-1711 (42kDa). Precision Plus Protein standards (BioRad) are annotated. B. Radiometric histone methyltransferaes assays comparing the catalytic activity of the wild type and R1625C variant when chicken olido-nucleosomes were the substrate. The amount of 3H-methyl incorporated is quantified as counts per minute (CPM) and error bars represent the standard error of the mean (n=3). A reaction without enzyme served as a negative control. C. Circular Diochroism (CD) absorbance spectra (plotted as the molar ellipticity ([Θ]) as a function of wavelength) comparing the secondary structure of wild type SETD2 (black) and the R1625C variant (purple). D. Thermal melt curves showing the change in CD absorbance at 207nm over the temperature range from 20-95°C for wild type SETD2 (black) and the R1625C variant (purple). E. Structural analysis of R1625. The crystal structure of the SETD2 SET domain (show in tan) bound to S-adenosyl-L-homocystein (SAH, shown in green) near the active site. Hydrogen bonds are shown as gray dashed lines (PDB code 4H12). F. In silico mutagenesis analysis (performed in PyMOL, Schrodinger Inc.)). The distances between the R1625C thiol and the carbonyl oxygens of A1617 and T1618 were measured in PyMOL (yellow dashed lines). G. Peptide pull-down assays comparing the binding of the wild type and the R1625C variant to the indicated histone H3 peptides. All peptides were biotinylated at the C-terminus and were either unmodified, or mono- (me1), di- (me2) or trimethylated (me3) at K36. Streptavidin coated magnetic beads without peptide served as the negative control.

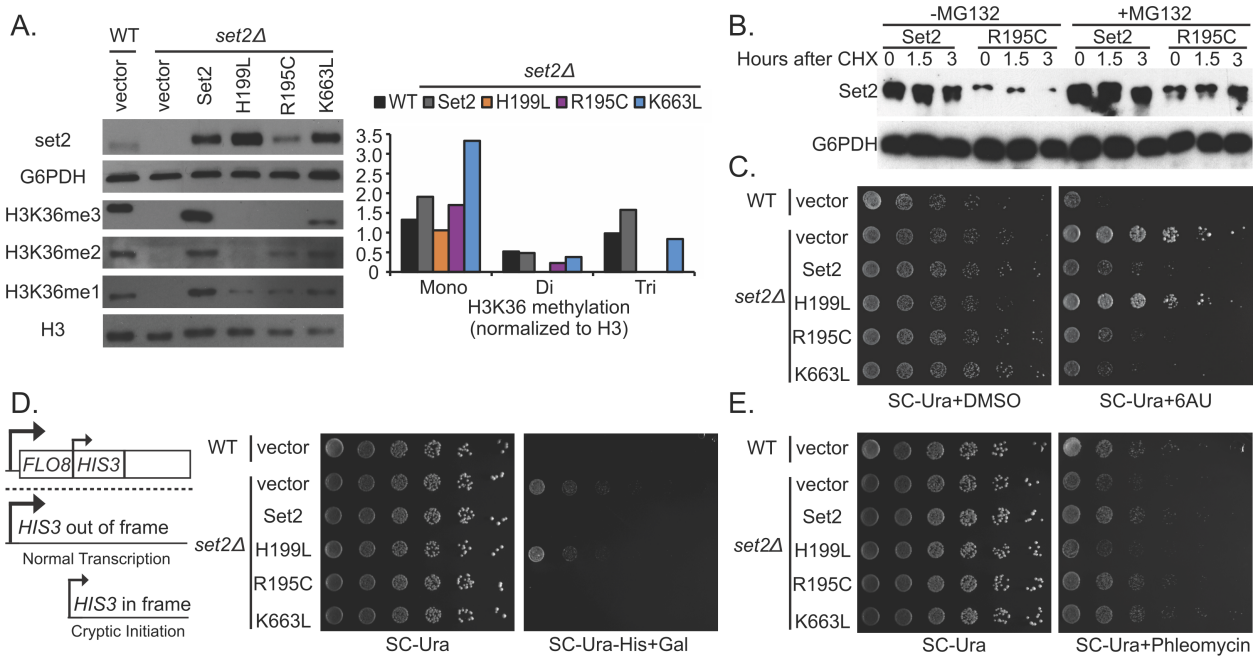


Figure A.4: Modeling of ccRCC specific mutations in Set2 results in separate effects based on H3K36me status

A. Anti-H3K36me immunoblots displaying levels of methylation in set2Δ yeast cells, as well as yeast with the indicated Set2 mutation. Quantification of H3K36me3/H3 is shown as a bar graph. B. Western results for Set2 and R195C protein levels after treatment with cyclohexamide (100μg/mL) and MG132 (75μM). C. 6-Azauracil (6-AU) treatment of wild-type or set2Δ yeast cells expressing the indicated Set2 mutations. D. Phleomycin treatment of wild-type or set2Δ yeast cells expressing the indicated Set2 mutations. E. Cryptic initiation assay of wild-type or set2Δ yeast cells expressing the indicated Set2 mutations.

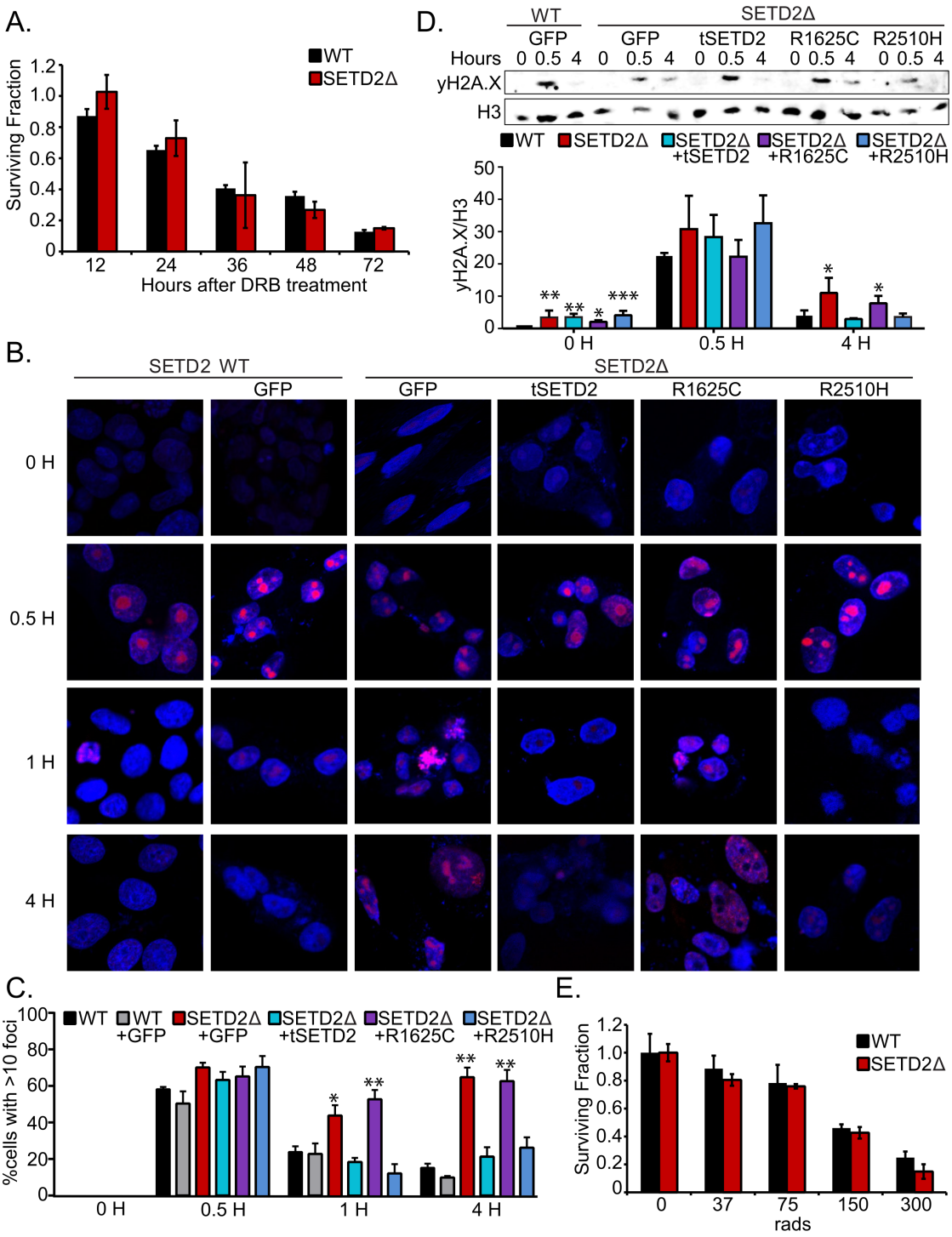

A.
Surviving Fraction
1.2
1.0
0.8
0.6
0.4
0.2
0
12
24
36
48
72
Hours after DRB treatment
WT
SETD2Δ
D.
WT
SETD2Δ
GFP
GFP
tSETD2
R1625C
R2510H
Hours 0 0.5 4 0 0.5 4 0 0.5 4 0 0.5 4 0 0.5 4
γH2A.X
H3
WT
SETD2Δ
SETD2Δ +tSETD2
SETD2Δ +R1625C
SETD2Δ +R2510H
γH2A.X/H3
50
40
30
20
10
0
0 H
0.5 H
4 H
B.
SETD2 WT
SETD2Δ
GFP
GFP
tSETD2
R1625C
R2510H
0 H
0.5 H
1 H
4 H
C.
%cells with >10 foci
100
80
60
40
20
0
WT
WT +GFP
SETD2Δ +GFP
SETD2Δ +tSETD2
SETD2Δ +R1625C
SETD2Δ +R2510H
0 H
0.5 H
1 H
4 H
E.
Surviving Fraction
1.2
1.0
0.8
0.6
0.4
0.2
0
0
37
75
150
300
rads
WT
SETD2Δ

Figure A.5: H3K36me3 loss delays γH2A.X foci resolution after DNA damage but does not alter viability

A. Surviving fraction of cells at 12, 24, 36, 48, and 72 hours post treatment with 100 μM 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB). Fraction was determined compared to an untreated control. Error bars represent standard deviation of triplicate treatments. B. γH2AX foci formation at 0, 0.5, 1 and 4 hours after irradiation (2Gy) in HKC wild-type or SETD2-inactivated cells transfected with GFP, tSETD2, R1625C mutant or R2510H mutant. The nuclei were visualized by DAPI staining. Representative immunofluorescence images are shown; scale bar at 10 μm. At least 5 fields were taken from each condition and four independent experiments were performed. C. Percentage of HKC cells with more than 10 γH2AX foci per cell at 0, 0.5, 1, and 4 hours after irradiation (2Gy). Error bars represent standard error. *$p<0.05$, **$p<0.01$, (two-sided t-test, comparison to HKC wild-type). D. Immunoblot analysis for the expression of γH2AX and H3 (loading control) from the 293 wild-type or SETD2-inactivated cells transfected with GFP, tSETD2, SET domain R1625C mutant or SRI domain R2510H mutant. The cells were irradiated by 2 Gy, and histones were acid-extracted at various time points. Average quantification of γH2AX/H3 after irradiation in three independent western blots Error bars represent standard error. *$p<0.05$, **$p<0.01$, ***$p<0.001$ (two-sided t-test, comparison to 293 wild-type + GFP). E. Radiation foci formation assay. Surviving fraction represents ratio of treatment (37, 75, 150, 300 rads) to 0 rad comparison. Error bars represent standard deviation of triplicate results.

REFERENCES

Ahn, S.H., Keogh, M.C., and Buratowski, S. (2009). Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. EMBO J. 28, 205–212.

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.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and Methylation of Histones and their Possible Role in the Regulation of RNA Synthesis. PNAS 315, 786–794.

Andersen, E.C., and Horvitz, H.R. (2007). Two *C. elegans* histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. Development 2999, 2991–2999.

Ares, M., Grate, L., and Pauling, M.H. (1999). A handful of intron-containing genes produces the lion ' s share of yeast mRNA. RNA 5, 1138–1139.

Awwad, S.W., and Ayoub, N. (2015). Overexpression of KDM4 lysine demethylases disrupts the integrity of the DNA mismatch repair pathway. Biol. Open 4, 498–504.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P., et al. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat. Struct. Mol. Biol. 21, 366–374.

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

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243–247.

Beaver, J.E., and Waters, M.L. (2016). Molecular Recognition of Lys and Arg Methylation. Chem. Biol. 11, 643–653.

Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N.D., 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.

Bender, L.B., Suh, J., Carroll, C.R., Fong, Y., Fingerman, I.M., Briggs, S.D., Cao, R., Zhang, Y., Reinke, V., and Strome, S. (2006). MES-4 : an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the *C. elegans* germ line. Development 3917, 3907–3917.

Biggar, K.K., and Li, S.S. (2015). Non-histone protein methylation as a regulator of cellular signalling and function. Nat. Rev. Mol. Cell Biol. 16.

Biggar, K.K., Wang, Z., and Li, S.S. (2017). SnapShot : Lysine Methylation beyond Histones. Mol. Cell 68, 1016-1016.e1.

Bilokapic, S., and Halic, M. (2019). Nucleosome and ubiquitin position Set2 to methylate H3K36. Nat. Commun. 10, 1–9.

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. EMBO J. 25, 4479–4489.

Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. Mol. Cell 48, 491–507.

Bortvin, A., and Winston, F. (1996). Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones. Science (80-.). 272, 1473–1476.

Bradbury, E.M. (1992). Reversible Histone Modifications and the Chromsome Cell Cycle. BioEssays 14, 9–16.

Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001). Coiled coils: A highly versatile protein folding motif. Trends Cell Biol. 11, 82–88.

Buuh, Z.Y., Lyu, Z., and Wang, R.E. (2017). Interrogating the Roles of Post-Translational Modifications of Non-Histone Proteins. J. Med. Chem. 61, 3239–3252.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., et al. (2013). An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting. Mol. Cell 49, 571–582.

Cancer Genome Atlas Research Network (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499, 43–49.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 507, 315–322.

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

Carvalho, S., Raposo, A.C., Martins, F.B., Grosso, A.R., Sridhara, S.C., Rino, J., Carmo-Fonseca, M., and de Almeida, S.F. (2013). Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. Nucleic Acids Res. 41, 2881–2893.

Carvalho, S., Vítor, A.C., Sridhara, S.C., Martins, F.B., Raposo, A.C., Desterro, J.M.P., Ferreira, J., and Almeida, S.F. de (2014). SETD2 is required for DNA double-strand break repair and activation of the p53- mediated checkpoint. Elife 3, 1–19.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2, 401–404.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. a, Somia, N. V, Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 39, e82.

Cheng, X., Collins, R.E., and Zhang, X. (2005). Structural and Sequence Motifs of Protein (Histone) Methylation Enzymes. Annu Rev Biophys Biomol Struct 24, 267–294.

Cheng, X., Jobin-robitaille, O., Billon, P., Buisson, R., Niu, H., Lacoste, N., Abshiru, N., Cote, V., Thibault, P., Kron, S.J., et al. (2018). Phospho-dependent recruitment of the yeast NuA4 acetyltransferase complex by MRX at DNA breaks regulates RPA dynamics during resection. Proc. Natl. Acad. Sci. 115, 1–6.

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.

Chin, H.G., Estève, P.-O., Pradhan, M., Benner, J., Patnaik, D., Carey, M.F., and Pradhan, S. (2007). Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. Nucleic Acids Res. 35, 7313–7323.

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 The Bur1 Cyclin-Dependent Protein Kinase Is Required for the Normal Pattern of Histone Methylation by Set2. Mol. Cell. Biol. 26, 3029–3038.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368–373.

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. J. Biol. Chem. 280, 5563–5570.

Corsi, A.K., Wightman, B., and Chalfi, M. (2015). A Transparent Window into Biology : A Primer on *Caenorhabditis elegans*. Genetics 200, 387–407.
Cosgrove, M.S., Boeke, J.D., and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. Nat. Struct. Mol. Biol. 11, 1037–1043.

Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., et al. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 463, 360–363.

Daugaard, M., Baude, A., Fugger, K., Povlsen, L.K., Beck, H., Sørensen, C.S., Petersen, N.H.T., Sorensen, P.H.B., Lukas, C., Bartek, J., et al. (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. Nat. Struct. Mol. Biol. 19, 803–810.

De Castro, M. (2016). Johann Gregor Mendel : paragon of experimental science. Mol. Genet. Genomic Med. 3–8.

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.

Diebold, M., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010). Noncanonical Tandem SH2 Enables Interaction of Elongation Factor Spt6 with RNA Polymerase II. J. Biol. Chem. 285, 38389–38398.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol. 6, 227.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Dominguez, D., Tsai, Y., Gomez, N., Jha, D.K., Davis, I., and Wang, Z. (2016). A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. Cell Res. 26, 946–962.

Dronamraju, R., and Strahl, B.D. (2014). A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. Nucleic Acids Res. 42, 870–881.

Dronamraju, R., Jha, D.K., Eser, U., Adams, A.T., Dominguez, D., Choudhury, R., Chiang, Y.C., Rathmell, W.K., Emanuele, M.J., Churchman, L.S., et al. (2017). Set2 methyltransferase facilitates cell cycle progression by maintaining transcriptional fidelity. Nucleic Acids Res. 46, 1331–1344.

Dronamraju, R., Kerschner, J.L., Peck, S.A., Hepperla, A.J., Adams, A.T., Hughes, K.D., Aslam, S., Yoblinski, A.R., Davis, I.J., Mosley, A.L., et al. (2018). Casein Kinase II Phosphorylation of Spt6 Enforces Article Casein Kinase II Phosphorylation of Spt6 Enforces Transcriptional Fidelity by Maintaining Spn1-Spt6 Interaction. Cell Rep. 25, 3476–3489.

Drouin, S., Laramée, L., Jacques, P.É., Forest, A., Bergeron, M., and Robert, F. (2010). DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. PLoS Genet. 6, 1–12.

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.

Duan, G., and Walther, D. (2015). The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. PLoS Comput. Biol. 11, e1004049.

Duns, G., Hofstra, R.M.W., Sietzema, J.G., Hollema, H., Duivenbode, I. Van, Kuik, A., Giezen, C., Osinga, J., Bergsma, J.J., Bijnen, H., et al. (2012). Targeted Exome Sequencing in Clear Cell Renal Cell Carcinoma Tumors Suggests Aberrant Chromatin Regulation as a Crucial Step in ccRCC Development. Hum. Mutat. 33, 1059–1062.

Duns, G., van den Berg, E., van Duivenbode, I., Osinga, J., Hollema, H., Hofstra, R.M.W., and Kok, K. (2010). Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. Cancer Res. 70, 4287–4291.

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.

Faber, P.W., Barnes, G.T., Srinidhi, J., Chen, J., Gusella, J.F., and Macdonald, M.E. (1998). Huntingtin interacts with a family of WW domain proteins. Hum. Mol. Genet. 7, 1463–1474.

Fang, D., Gan, H., Lee, J.-H., Han, J., Wang, Z., Riester, S.M., Jin, L., Chen, J., Zhou, H., Wang, J., et al. (2016). The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science (80-.). 352, 1344–1348.

Ficarro, S.B., McCleland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., and White, F.M. (2002). Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae* . Nat. Biotechnol. 20, 301–305.

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.

Fontebasso, A.M., Schwartzentruber, J., Khuong-Quang, D.-A., Liu, X.-Y., Sturm, D., Korshunov, A., Jones, D.T.W., 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.

Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., and Stratton, M.R. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). In Current Protocols in Human Genetics, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p.

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. J. Biol. Chem. 287, 3249–3256.

Furuhashi, H., Takasaki, T., Rechtsteiner, A., Li, T., Kimura, H., Checchi, P.M., Strome, S., and Kelly, W.G. (2010). Trans-generational epigenetic regulation of C . elegans primordial germ cells. Epigenetics Chromatin 3, 1–21.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci. Signal. 6, pl1–pl1.

Gardner, K.E., Allis, C.D., and Strahl, B.D. (2011). OPERating ON Chromatin , a Colorful Language where Context Matters. J. Mol. Biol. 409, 36–46.

Gayon, J. (2016). From Mendel to epigenetics: History of genetics. C. R. Biol. 339, 225–230.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. N. Engl. J. Med. 366, 883–892.

Gilbert, T.M., McDaniel, S.L., Byrum, S.D., Cades, J. a., Dancy, B.C.R., Wade, H., Tackett, a. J., Strahl, B.D., and Taverna, S.D. (2014). A PWWP Domain-Containing Protein Targets the

NuA3 Acetyltransferase Complex via Histone H3 Lysine 36 trimethylation to Coordinate Transcriptional Elongation at Coding Regions. Mol. Cell. Proteomics 13, 2883–2895.

Gopalakrishnan, R., Marr, S.K., Kingston, R.E., and Winston, F. (2019). A conserved genetic interaction between Spt6 and Set2 regulates H3K36 methylation. Nucleic Acids Res. 47, 1–16.

Gouw, M., Michael, S., Samano-Sanchez, H., Kumar, M., Zeke, A., Lang, B., Bely, B., Chemes, L.B., Davey, N.E., Deng, Z., et al. (2018). The eukaryotic linear motif resource – 2018 update. Nucleic Acids Res. 46, 428–434.

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.

Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vítor, A.C., Desterro, J.M.P., Carmo-fonseca, M., and Almeida, S.F. De (2015). Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal. Elife 4, 1-16.

Guo, R., Zheng, L., Park, J.W., Lv, R., Chen, H., Jiao, F., Xu, W., Mu, S., Wen, H., Qiu, J., et al. (2014). BS69/ZMYND11 Reads and Connects Histone H3.3 Lysine 36 Trimethylation-Decorated Chromatin to Regulated Pre-mRNA Processing. Mol. Cell 1–13.

Habibian, J., and Ferguson, B.S. (2019). The Crosstalk between Acetylation and Phosphorylation: Emerging New Roles for HDAC Inhibitors in the Heart. Int. J. Mol. Sci. 20, 102.

Hacker, K.E., Fahey, C.C., Shinsky, S.A., Chiang, Y.-C.J., DiFiore, J. V., Jha, D.K., Vo, A.H., Shavit, J.A., Davis, I.J., Strahl, B.D., et al. (2016). Structure/Function Analysis of Recurrent Mutations in SETD2 Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and H3K36 Trimethylation. J. Biol. Chem. 291, jbc.M116.739375.

Hales, K.G., Korey, C.A., Larracuente, A.M., and Roberts, D.M. (2015). Genetics on the Fly : A Primer on the Drosophila Model System. Genetics 201, 815–842.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell 38, 576–589.

Hoffman, C.S., Wood, V., and Fantes, P.A. (2015). An ancient yeast for young geneticists: A primer on the Schizosaccharomyces pombe model system. Genetics 201, 403–423.

Holt, L.J., Tuch, B.B., Villen, J., Johnson, a. D., Gygi, S.P., and Morgan, D.O. (2009). Global Analysis of Cdk1 Substrate Phosphorylation Sites Provides Insights into Evolution. Science (80-.). 325, 1682–1686.

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. 107, 2956–2961.

Hu, Z., Chen, K., Xia, Z., Chavez, M., Pal, S., Seol, J.-H., Chen, C.-C., Li, W., and Tyler, J.K. (2014). Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. Genes Dev. 28, 396–408.

Huber, F., Bunina, D., Gupta, I., Theer, P., Steinmetz, L.M., and Knop, M. (2016). Protein Abundance Control by Non-coding Antisense Transcription. CellReports 15, 1–12.
Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp. Mol. Med. 49, e324-22.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., et al. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast 21, 947–962.

Jeronimo, C., Watanabe, S., Kaplan, C.D., Peterson, C.L., and Robert, F. (2015). The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. Mol. Cell 58, 1113–1123.

Jha, D.K., and Strahl, B.D. (2014). An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair. Nat. Commun. 5.

Ji, Z., Sheng, Y., Miao, J., Li, X., Zhao, H., Wang, J., Cheng, C., Wang, X., Liu, K., Zhang, K., et al. (2019). The histone methyltransferase Setd2 is indispensable for V(D)J recombination. Nat. Commun. 10, 1–14.

Jiang, C., He, C., Wu, Z., Li, F., and Xiao, J. (2018). Histone methyltransferase SETD2 regulates osteosarcoma cell growth and chemosensitivity by suppressing Wnt/beta-catenin signaling. Biochem. Biophys. Res. Commun. 502, 382–388.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to pol II elongation. Mol. Cell 20, 971–978.

Kanu, N., Grönroos, E., Martinez, P., Burrell, R. a, Yi Goh, X., Bartkova, J., Maya-Mendoza, a, Mistrík, M., Rowan, a J., Patel, H., et al. (2015). SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. Oncogene 1–10.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription Elongation Factors Repress Transcription Inititaion from Cryptic Sites. Science 239, 17–18.

Keogh, M.C., Kim, J.A., Downey, M., Fillingham, J., Chowdhury, D., Harrison, J.C., Onishi, M., Datta, N., Galicia, S., Emili, A., et al. (2006a). A phosphatase complex that dephosphorylates γH2AX regulates DNA damage checkpoint recovery. Nature 439, 497–501.

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

Keogh, M.C., Kim, J.A., Downey, M., Fillingham, J., Chowdhury, D., Harrison, J.C., Onishi, M., Datta, N., Galicia, S., Emili, A., et al. (2006a). A phosphatase complex that dephosphorylates γH2AX regulates DNA damage checkpoint recovery. Nature 439, 497–501.

Keogh, M.C., Mennella, T.A., Sawa, C., Berthelet, S., Krogan, N.J., Wolek, A., Podolny, V., Carpenter, L.R., Greenblatt, J.F., Baetz, K., et al. (2006b). The *Saccharomyces cerevisiae* histone H2A variant Htz1 is acetylated by NuA4. Genes Dev. 20, 660–665.

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

Kim, H.S., Rhee, D.K., and Jang, Y.K. (2008). Methylations of histone H3 lysine 9 and lysine 36 are functionally linked to DNA replication checkpoint control in fission yeast. Biochem. Biophys. Res. Commun. 368, 419–425.

Kim, J.H., Lee, B.B., Oh, Y.M., Zhu, C., Steinmetz, L.M., Lee, Y., Kim, W.K., Lee, S.B., Buratowski, S., and Kim, T. (2016). Modulation of mRNA and lncRNA expression dynamics by the Set2–Rpd3S pathway. Nat. Commun. 7, 13534.

Kim, T., Xu, Z., Clauder-Munster, S., Steinmetz, L.M., and Buratowski, S. (2012). Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. Cell 150, 1158–1169.

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. Mol. Cell. Biol. 25, 3305–3316.

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, 376–382.

Kreher, J., Takasaki, T., Cockrum, C., Sidoli, S., Garcia, B.A., Jensen, O.N., and Strome, S. (2018). Distinct Roles of Two Histone Methyltransferases in Transmitting H3K36me3-Based Epigenetic Memory Across Generations in *Caenorhabditis elegans*. Genetics 210, 969–982.

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.

Lee, C.H., Wu, J., and Li, B. (2013). Chromatin remodelers fine-tune h3k36me-directed deacetylation of neighbor nucleosomes by Rpd3S. Mol. Cell 52, 255–263.

Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A.C.H., Margaritis, T., van de Pasch, L.A.L., van Heesch, S.A.A.C., Brok, M.O., Groot Koerkamp, M.J.A., et al. (2011). The Specificity and Topology of Chromatin Interaction Pathways in Yeast. Mol. Cell 42, 536–549.

Leung, C.S., Douglass, S.M., Morselli, M., Obusan, M.B., Pavlyukov, M.S., Pellegrini, M., Johnson, T.L., Leung, C.S., Douglass, S.M., Morselli, M., et al. (2019). H3K36 Methylation and the Chromodomain Protein Eaf3 Are Required for Proper Cotranscriptional Spliceosome Assembly Cell Reports 27, 3760–3769.e4.

Li, B., Gogol, M., Carey, M., Lee, D., Seidel, C., and Workman, J.L. (2007). Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. Science 316, 1050–1054.

Li, B., Howe, L.A., Anderson, S., Yates, J.R., and Workman, J.L. (2003). The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. J. Biol. Chem. 278, 8897–8903.

Li, B., Jackson, J., Simon, M.D., Fleharty, B., Gogol, M., Seidel, C., Workman, J.L., and Shilatifard, A. (2009). 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.

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 MutSa. Cell 153, 590–600.

Li, J., Duns, G., Westers, H., Sijmons, R., and Berg, A. Van Den (2016). SETD2 : an epigenetic functionality modifier with tumor suppressor. Oncotarget 7.

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. Proc. Natl. Acad. Sci. U. S. A. 102, 17636–17641.

Lickwar, C.R., Rao, B., Shabalin, A.A., Nobel, A.B., Strahl, B.D., and Lieb, J.D. (2009). The set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency. PLoS One 4.

Lu, C., Jain, S.U., Hoelper, D., Bechet, D., Molden, R.C., Ran, L., Murphy, D., Venneti, S., Hameed, M., Pawel, B.R., et al. (2016). Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science 352, 844–849.

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

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, 996–1001.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260.

Mahadevan, L.C., Willis, A.C., and Barratt, M.J. (1991). Rapid Histone H3 Phosphorylation in Response to Growth Factors, Phorbol Esters Okadaic Acid, and Protein Synthesis Inhibitors. Cell 65, 775–783.

Maltby, V.E., Martin, B.J.E., Schulze, J.M., Johnson, I., Hentrich, T., Sharma, A., Kobor, M.S., and Howe, L. (2012). Histone H3 lysine 36 methylation targets the Isw1b remodeling complex to chromatin. Mol. Cell. Biol. 32, 3479–3485.

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. Nat. Commun. 5, 3469.

Mar, B.G., Chu, S.H., Kahn, J.D., Krivtsov, A. V., Koche, R., Castellano, C.A., Kotlier, J.L., Zon, R.L., McConkey, M.E., Chabon, J., et al. (2017). SETD2 alterations impair DNA damage recognition and lead to resistance to chemotherapy in leukemia. Blood 130, 2631–2641.

Martin, B.J.E., McBurney, K.L., Maltby, V.E., Jensen, K.N., Brind'Amour, J., and Howe, L. (2017). Histone H3K4 and H3K36 Methylation Independently Recruit the NuA3 Histone Acetyltransferase in *Saccharomyces cerevisiae* . Genetics 205, genetics.116.199422.

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The Abundant Histone Chaperones Spt6 and FACT Collaborate to Assemble, Inspect, and Maintain Chromatin Structure in *Saccharomyces cerevisiae* . Genetics 201, 1031–1045.

McDaniel, S.L., and Strahl, B.D. (2017). Shaping the cellular landscape with Set2 / SETD2 methylation. Cell. Mol. Life Sci. 74, 3317–3334.

McDaniel, S.L., Fligor, J.E., Ruan, C., Cui, H., Bridgers, J.B., DiFiore, J. V., Guo, A.H., Li, B., and Strahl, B.D. (2016). Combinatorial histone readout by the dual plant homeodomain (PHD) fingers of Rco1 mediates Rpd3S chromatin recruitment and the maintenance of transcriptional fidelity. J. Biol. Chem. 291, 14796–14802.

McDaniel, S.L., Hepperla, A.J., Huang, J., Dronamraju, R., Adams, A.T., Kulkarni, V.G., Davis, I.J., and Strahl, B.D. (2017). H3K36 Methylation Regulates Nutrient Stress Response in *Saccharomyces cerevisiae* by Enforcing Article H3K36 Methylation Regulates Nutrient Stress Response in *Saccharomyces cerevisiae* by Enforcing Transcriptional Fidelity. CellReports 19, 2371–2382.

McKay, D.J., Klusza, S., Penke, T.J., Meers, M.P., Curry, K.P., McDaniel, S.L., Malek, P.Y., Cooper, S.W., Tatomer, D.C., Lieb, J.D., et al. (2015). Interrogating the Function of Metazoan Histones using Engineered Gene Clusters. Dev Cell 32, 373–386.

Meerbrey, K.L., Hu, G., Kessler, J.D., Roarty, K., Li, M.Z., Fang, J.E., Herschkowitz, J.I., Burrows, A.E., Ciccia, A., Sun, T., et al. (2011). The pINDUCER lentiviral toolkit for inducible RNA interference *in vitro* and *in vivo*. Proc. Natl. Acad. Sci. U. S. A. 108, 3665–3670.

Meers, M.P., Henriques, T., Lavender, C.A., McKay, D.J., Strahl, B.D., Duronio, R.J., Adelman, K., and Matera, A.G. (2017). Histone gene replacement reveals a posttranscriptional role for H3K36 in maintaining metazoan transcriptome fidelity. Elife 6, 1–23.

Mendel, G. (1901). Experiments in Hybrdisation.pdf. J. R. Hortic. Soc.

Merker, J.D., Dominska, M., Greenwell, P.W., Rinella, E., Bouck, D.C., Shibata, Y., Strahl, B.D., Mieczkowski, P., and Petes, T.D. (2008). The histone methylase Set2p and the histone deacetylase Rpd3p repress meiotic recombination at the HIS4 meiotic recombination hotspot in *Saccharomyces cerevisiae* . DNA Repair (Amst). 7, 1298–1308.

Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., Ottonello, S., et al. (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife 1, 1–21.

Narita, T., Weinert, B.T., and Choudhary, C. (2019). Functions and mechanisms of non-histone protein acetylation. Nat. Rev. Mol. Cell Biol. 20.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., Jacquier, A., Aubenton-carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457, 1038–1042.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature 543, 72–77.

Nielsen, M., Ard, R., Leng, X., Ivanov, M., Kindgren, P., Pelechano, V., and Marquardt, S. (2019). Transcription-driven chromatin repression of Intragenic transcription start sites. PLoS Genet. 15, 1–33.

Oughtred, R., Stark, C., Breitkreutz, B., Rust, J., Boucher, L., Chang, C., Kolas, N., Donnell, L.O., Leung, G., Mcadam, R., et al. (2018). The BioGRID interaction database : 2019 update. Nucleic Acids Res. 47, 529–541.

Pai, C.-C., Kishkevich, A., Deegan, R.S., Keszthelyi, A., Folkes, L., Kearsey, S.E., De Leon, N., Soriano, I., de Bruin, R.A.M., Carr, A.M., et al. (2017). Set2 Methyltransferase Facilitates DNA Replication and Promotes Genotoxic Stress Responses through MBF-Dependent Transcription. Cell Rep. 20, 2693–2705.

Pai, 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. Nat. Commun. 5.

Papamichos-Chronakis, M., and Peterson, C.L. (2013). Chromatin and the genome integrity network. Nat. Rev. Genet. 14, 62–75.

Papillon-Cavanagh, S., Lu, C., Gayden, T., Mikael, L.G., Bechet, D., Karamboulas, C., Ailles, L., Karamchandani, J., Marchione, D.M., Garcia, B.A., et al. (2017). Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat. Genet. 49, 180–185.
Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.-C., Davis, I.J., Fahey, C.C., Hacker, K.E., et al. (2016). Dual Chromatin and Cytoskeletal Remodeling by SETD2. Cell 166, 950–962.

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, 1–10.

Pawson, T., and Scott, J.D. (2005). Protein phosphorylation in signaling – 50 years and counting. Trends Biochem. Sci. 30, 286–290.

Pfister, S.X., Ahrabi, S., Zalmas, L.P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday, T., Legube, G., LaThangue, N.B., Porter, A.C.G., et al. (2014). SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. Cell Rep. 7, 2006–2018.

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.

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 Genet. 8, e1002717.

Pryde, F., Jain, D., Kerr, A., Curley, R., Mariotti, F.R., and Vogelauer, M. (2009). H3 K36 methylation helps determine the timing of Cdc45 association with replication origins. PLoS One 4.

Pu, M., Ni, Z., Wang, M., Wang, X., Wood, J.G., Helfand, S.L., Yu, H., and Lee, S.S. (2015). Trimethylation of Lys36 on H3 restricts gene expression change during aging and impacts life span. Genes Dev. 29, 718–731.

Quan, T.K., and Hartzog, G.A. (2009). Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of *Saccharomyces cerevisiae* Spt4-Spt5 in transcription. Genetics 184, 321–334.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Racusen, L.C., Monteil, C., Sgrignoli, A., Lucskay, M., Marouillat, S., Rhim, J.G.S., and Morin, J.-P.P. (1997). Cell lines with extended *in vitro* growth potential from human renal proximal tubule: Characterization, response to inducers, and comparison with established cell lines. J. Lab. Clin. Med. 129, 318–329.

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. Mol. Cell. Biol. 25, 9447–9459.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specificc histone H3 methyltransferases. Nature 406, 593–599.

Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T.M., Egelhofer, T.A., Wang, W., Kimura, H., Lieb, J.D., and Strome, S. (2010). The Histone H3K36 Methyltransferase MES-4 Acts Epigenetically to Transmit the Memory of Germline Gene Expression to Progeny. PLoS Genet. 6.

Rona, G.B., Eleutherio, E.C.A., and Pinheiro, A.S. (2016). PWWP domains and their modes of sensing DNA and histone methylated lysines. Biophys. Rev. 8, 63–74.

Rondelet, G., Dal Maso, T., Willems, L., and Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J. Struct. Biol. 194, 357–367.

Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. Biochim. Biophys. Acta - Gene Regul. Mech. 1839, 627–643.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738.

Ruan, C., Lee, C.H., Cui, H., Li, S., and Li, B. (2015). Nucleosome contact triggers conformational changes of Rpd3S driving high-affinity H3K36me nucleosome engagement. Cell Rep. 10, 204–215.

Sampath, S.C., Marazzi, I., Yap, K.L., Sampath, S.C., Krutchinsky, A.N., Viale, A., Rudensky, E., Zhou, M., Chait, B.T., and Tarakhovsky, A. (2007). Methylation of a Histone Mimic within the Histone Methyltransferase G9a Regulates Protein Complex Assembly. Mol. Cel 27, 596–608.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.-R.J. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat. Biotechnol. 29, 697–698.

Schrödinger, LLC (2015). The {PyMOL} Molecular Graphics System, Version~1.8.
Sen, P., Dang, W., Donahue, G., Dai, J., Dorsey, J., Cao, X., Liu, W., Cao, K., Perry, R., Lee, J.Y., et al. (2015). H3K36 methylation promotes longevity by enhancing transcriptional fidelity. Genes Dev. 29, 1362–1376.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.-H.A., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G.E., Fingerman, I.M., Briggs, S.D., et al. (2007). Proteome-wide Analysis in *Saccharomyces cerevisiae* Identifies Several PHD Fingers as Novel Direct and Selective Binding Modules of Histone H3 Methylated at Either Lysine 4 or Lysine 36. J. Biol. Chem. 282, 2450–2455.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

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. J. Biol. Chem. 287, 1709–1718.

Simon, J.M., Hacker, K.E., Singh, D., Brannon, A.R., Parker, J.S., Weiser, M., Ho, T.H., Kuan, P.-F., Jonasch, E., Furey, T.S., et al. (2014). Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res. 24, 241–250.

Smolle, M., and Workman, J.L. (2013). Transcription-associated histone modifications and cryptic transcription. Biochim. Biophys. Acta - Gene Regul. Mech. 1829, 84–97.

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. Nat. Struct. Mol. Biol. 19, 884–892.

Sorenson, M.R., Jha, D.K., Ucles, S. a, Flood, D.M., Strahl, B.D., Stevens, S.W., and Kress, T.L. (2016). Histone H3K36 methylation regulates pre-mRNA splicing in *Saccharomyces cerevisiae* . RNA Biol. 13, 412–426.

Soshnev, A.A., Josefowicz, S.Z., and Allis, C.D. (2016). Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. Mol. Cell 62, 681–694.

Stabell, M., Larsson, J., Aalen, R.B., and Lambertsson, A. (2007). Drosophila dSet2 functions in H3-K36 methylation and is required for development. Biochem. Biophys. Res. Commun. 359, 784–789.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature 403, 41–45.

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. Mol. Cell. Biol. 22, 1298–1306.

Strahl, B.D., Ohba, R., Cook, R.G., and Allis, C.D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc. Natl. Acad. Sci. 96, 14967–14972.

Stuckey, S., and Storici, F. (2013). Gene Knockouts, *in vivo* Site-Directed Mutagenesis and Other Modifications Using the Delitto Perfetto System in *Saccharomyces cerevisiae* . Methods Enzymol. 533, 103–131.
Sudol, M., Chen, H.I., Bougeret, C., Einbond, A., and Bork, P. (1995). Characterization of a novel protein-binding module - the W W domain. FEBS Lett. 369, 67–71.

Sun, M., Larivie, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A Tandem SH2 Domain in Transcription Elongation Factor Spt6 Binds the Phosphorylated RNA Polymerase II C-terminal Repeat Domain (CTD). J. Biol. Chem. 285, 41597–41603.

Sun, S., Yang, F., Tan, G., Costanzo, M., Hirschman, J., Theesfeld, C., Bansal, P., Yi, S., Yu, A., Tyagi, T., et al. (2016). An extended set of yeast-based functional assays accurately identifies human disease mutations. 670–680.

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. J. Biol. Chem. 280, 35261–35271.

Suzuki, S., Kato, H., Suzuki, Y., Chikashige, Y., Hiraoka, Y., Kimura, H., Nagao, K., Obuse, C., Takahata, S., and Murakami, Y. (2016). Histone H3K36 trimethylation is essential for multiple silencing mechanisms in fission yeast. Nucleic Acids Res. 7, 1–16.

Thompson, D.M., and Parker, R. (2007). Cytoplasmic Decay of Intergenic Transcripts in *Saccharomyces cerevisiae* . Mol. Cell. Biol. 27, 92–101.

Tsubota, T., Berndsen, C.E., Erkmann, J.A., Smith, C.L., Yang, L., Freitas, M.A., Denu, J.M., and Kaufman, P.D. (2007). Histone H3K56 Acetylation Is Catalyzed by Histone Chaperone-Dependent Complexes. Mol. Cell 25, 703–712.

Ubersax, J.A., and Ferrell, J.E. (2007). Mechanisms of specificity in protein phosphorylation. Nat. Rev. Mol. Cell Biol. 8, 530–542.

van Dijk, E.L., Chen, C.L., D'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature 475, 114–117.

VanDemark, A.P., Kasten, M.M., Ferris, E., Heroux, A., Hill, C.P., and Cairns, B.R. (2007). Autoregulation of the Rsc4 Tandem Bromodomain by Gcn5 Acetylation. Mol. Cell 27, 817–828.

Vary, J.C., Gangaraju, V.K., Qin, J., Landel, C.C., Kooperberg, C., Bartholomew, B., and Tsukiyama, T. (2003). Yeast Isw1p Forms Two Separable Complexes In Vivo. Mol. Cell. Biol. 23, 80–91.

Venkatesh, S., and Workman, J.L. (2013). Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. Wiley Interdiscip. Rev. Dev. Biol. 2, 685–700.

Venkatesh, S., Li, H., Gogol, M.M., and Workman, J.L. (2016). Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. Nat. Commun. 7, 13610.

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.

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. J. Biol. Chem. 281, 13–15.

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.

Wang, S., Zhou, B.O., and Zhou, J. (2011). Histone H3 Lysine 4 Hypermethylation Prevents Aberrant Nucleosome Remodeling at the PHO5 Promoter. Mol. Cell. Biol. 31, 3171–3181.

Wang, Y., Niu, Y., and Li, B. (2015). Balancing acts of SRI and an auto-inhibitory domain specify Set2 function at transcribed chromatin. Nucleic Acids Res. 43, 4881–4892.

Washburn, M.P., Wolters, D., and Yates, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247.

Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. Curr. Protoc. Bioinforma. 54, 1–55.

Wei, W., Hennig, B.P., Wang, J., Zhang, Y., Piazza, I., Sanchez, Y.P., Chabbert, C.D., Adjalley, S.H., Steinmetz, L.M., and Pelechano, V. (2019). Chromatin-sensitive cryptic promoters putatively drive expression of alternative protein isoforms in yeast. Genome Res. 12, 1974–1984.

Weiner, A., Hsieh, T.-H.S., Appleboim, A., Chen, H. V., Rahat, A., Amit, I., Rando, O.J., and Friedman, N. (2015). High-Resolution Chromatin Dynamics during a Yeast Stress Response. Mol. Cell 58, 371–386.

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.

Williams, S.K., Truong, D., and Tyler, J.K. (2008). Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. PNAS 105, 9000–9005.

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.

Wozniak, G.G., and Strahl, B.D. (2014). Hitting the “mark”: Interpreting lysine methylation in the context of active transcription. Biochim. Biophys. Acta - Gene Regul. Mech. 1839, 1353–1361.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121, 725–737.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. 17, 654–663.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature 457, 1033–1037.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2014). The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8.

Yang, J. and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 43, 174–181.

Yang, S., Zheng, X., Lu, C., Li, G., Allis, C.D., and Li, H. (2016). Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. Genes Dev. 30, 1611–1616.

Yap, K.L., and Zhou, M.-M. (2011). Structure and Mechanisms of Lysine Methylation Recognition by the Chromodomain in Gene Transcription. Biochemistry 50, 1966–1980.

Yoh, S.M., Lucas, J.S., and Jones, K. a (2008). The Iws1 : Spt6 : CTD complex controls cotranscriptional mRNA The Iws1 : Spt6 : CTD complex controls cotranscriptional mRNA biosynthesis and HYPB / Setd2-mediated histone H3K36 methylation. Genes Dev. 3422–3434.

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. Mol. Cell. Biol. 28, 4915–4926.

Yuan, H., Li, N., Fu, D., Ren, J., Hui, J., Peng, J., Liu, Y., Qiu, T., Jiang, M., Pan, Q., et al. (2017). Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. J. Clin. Invest. 127, 3375–3392.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res. 21, 564–578.

Zaborowska, J., Egloff, S., and Murphy, S. (2016). The pol II CTD: new twists in the tail. Nat Struct Mol Biol 23, 771–777.

Zbar, B., Brauch, H., Talmadge, C., and Linehan, M. (1987). Loss of alleles of loci on the short arm of chromosome 3 in renal cell carcinoma. Nature 327, 721–724.

Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S.L., Payne-Turner, D., Easton, J., Chen, X., Wang, J., Rusch, M., et al. (2012). The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 481, 157–163.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 1–8.

Zhang, Y., Shan, C., Wang, J., Bao, K., Tong, L., and Jia, S. (2017). Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. Sci. Rep. 1–9.

Zheng, W., Ibáñez, G., Wu, H., Blum, G., Zeng, H., Dong, A., Li, F., Hajian, T., Allali-Hassani, A., Amaya, M.F.M.., et al. (2012). Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. J.Am.Chem.Soc. 134, 18004–18014.

Zhu-Yr, Peery, T., Peng-Tm, Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews-Mb, and Price-Dh-(ReprintAuthor) (1997). Transcription elongation factor p tefb is required for hiv 1 tat transactivation *in vitro*. Genes Dev. 11, 2622–2632.

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.