THE EFFECT OF CANCER-ASSOCIATED SETD2 MUTATIONS ON TRANSCRIPTION AND CHROMATIN ORGANIZATION

Catherine C. Fahey

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.

Chapel Hill
2017

Approved by:

Albert S. Baldwin

Ian J. Davis

William Y. Kim

W. Kimryn Rathmell

Brian D. Strahl

© 2017
Catherine C. Fahey
ALL RIGHTS RESERVED

ABSTRACT

Catherine C. Fahey: The effect of cancer-associated SETD2 mutations on transcription and chromatin organization
(Under the direction of W. Kimryn Rathmell and Ian J. Davis)

Clear cell renal cell carcinoma is characterized by mutations in chromatin modifying enzymes. Among these is SETD2, a non-redundant histone H3 lysine 36 methyltransferase. Mutations in SETD2 in ccRCC are either early-inactivating, occur in the catalytic SET domain, or are found in the Set2 Rpb1-interacting domain. We inactivated SETD2 in ccRCC cells lines and reintroduced a truncated but functional wildtype SETD2 (tSETD2), as well as three ccRCC-associated point mutations in order to examine the effect on DNA damage repair, chromatin organization, transcription, and H3K36 trimethylation. We found that SETD2 loss results in complete loss of H3K36me3. SETD2Δ cells do not resolve γH2A.X foci after DNA damage and show altered chromatin accessibility. Transcription is largely unaffected by SETD2 loss. tSETD2 restores H3K36me3 to loci marked by H3K36me3 in wildtype cells, indicating that the N-terminus of SETD2 in not required for H3K36me3 placement. The first examined point mutant, R1625C, occurs in the SET domain and disrupts catalytic activity of SETD2 by reducing the interaction between SETD2 and histone H3, which results in decreased protein stability in the cell. The second examined point mutation, R2510H, occurs in the SRI domain. This SRI domain mutant is catalytically active. The third mutation, T2457*, deletes the SRI domain from tSETD2, and is also catalytically active. Both the R2510H and the T2457* mutations disrupt the interaction between SETD2 and RNA polymerase II. Surprisingly, both of these mutants also restore H3K36me3 to loci marked by wildtype SETD2. These mutants also show increased H3K36me3 levels near the TSS relative to wildtype. This suggests that the SRI domain is necessary for interaction with RNAPII, is not required for normal H3K36me3 placement across the genome, and may be involved at the TSS for H3K36me3 marking. The separation of function for the SET and SRI domains is of important consideration when developing therapeutics which target SETD2 mutation in ccRCC.

To my parents, Patrick and Carol Fahey. Thank you for everything.

ACKNOWLEDGMENTS

I am forever grateful for my fabulous mentors, Dr. W. Kimryn Rathmell and Dr. Ian Davis. Both are fantastic role models and represent MD/PhDs very well. I have had the opportunity to learn so much from both of them. Dr. Davis is an extremely intelligent, enthusiastic scientist whose passion and dedication to research inspires me daily. He is kind, generous, and his compassion for his patients is apparent to anyone who has the pleasure of working with him. Dr. Rathmell is a brilliant person, and her excitement for scientific discovery and dedication to mentorship motivates me to continue my journey towards my degrees. She is thoughtful, considerate, and supportive of all her mentees. Both of my mentors have provided scientific insight, career advice, and pushed me to develop as a young scientist. This dissertation would have been incomplete without their combined input, and I cannot thank them enough.

I would also like to thank the members of my thesis committee, Dr. Brian Strahl, Dr. Albert Baldwin and Dr. William Kim. I have had wonderful experiences meeting with my committee. Everyone asks good questions, is supportive of my ideas, and also very kind. Dr. Baldwin wrote a recommendation letter that was instrumental to the F30 I received from the NCI, and always asks about my research when I see him in the Lineberger Comprensive Cancer Center. He has been a fantastically supportive chair of my committee, and always asks in depth questions. Dr. Strahl has been heavily been involved in my project and his advice has been invaluable towards its completion. He has provided expertise for many of my experiments, and proposed new hypotheses that broadened the scope of this project. Dr. Kim was my longitudinal clerkship mentor, and I learned so much about clinical medicine from him. He has fantastic bedside manner, and learning from him was a wonderful experience. He also asks insightful questions in my meetings, and has followed up outside of meetings to check my progress. I have had a very supportive committee, both personally and scientifically, and I am very grateful.

I have been very fortunate to have a wonderful family who has encouraged me and put up with me during a very long process. My parents, Pat and Carol, never pushed me to be anything I did not want to

be. Instead, they expected me to try my best at whatever I wanted. When I decided to pursue an MD/PhD, this was still their goal. I know no matter what trajectory my career takes, they will be proud of me and continue to love and support me. Their love has been a wonderful blessing during the 5 years of my PhD. My older sister Kristin Weirich is a constant source of inspiration. I have looked up to Kristin my entire life. She is one of the smartest people I know, is kind, compassionate and very funny. It is easy to try to meet expectations when the person you look up to is constantly exceeding them. I would not trade my sister for anyone in the world. She married a wonderful man, Matt Weirich, whose presence has only added to our wonderful family. Matt is hilarious, a talented chef, and incredibly supportive of my sister. They recently had a daughter Clara Ruth (aka Smudgie) who is super cute. I am looking forward to watching her grow into a strong woman like her mother.

I am also grateful to the MD/PhD program and the Curriculum in Genetics and Molecular Biology. Both programs have provided opportunities to present my work and obtain feedback, and have been very responsive to student needs. In particular, the MD/PhD program has created a wonderful environment for physician-scientists in training that embraces the idea that everyone can succeed together. Dr. Deshmukh, Dr. Darville and the late Dr. Orringer have always gone out of their way to help the students in the program, and I am grateful for their guidance. I would also like to thank Alison Regan and Carol Herion for answering all of my questions, letting me come hang out in the MD/PhD office and for making sure I complete all of the paperwork that needs to be done. They have never let me fall through the cracks, and are wonderful kind people who run the MD/PhD program smoothly and efficiently.

I would also like to thank the people who supported me throughout the course of this PhD. I have a wonderful group of friends, but in particular I would like to acknowledge Jordan Walter for being available to get food, drinks and talk about the difficulties of grad school whenever I needed to. Mary Elizabeth Entwistle helped take care of my dog when I worked late nights in lab, and often made sure I ate. The ‘Monday Funday’ group has provided many nights of entertainment. Shannon Munchel has listened and asked questions even when she did not have any idea what I was talking about. Dr. Kate Gessner (née Hacker) has been a friend and mentor throughout my graduate career. She originated this project, and it would not have been possible without her support. I would also like to thank my boyfriend, Dr. Kenneth D. Stewart for the food, conversation, love and laughter of the last year and a half.

Finally, I would like to thank my lab mates. I really would not have been able to complete this dissertation without them. Dr. Yun-Chen ‘Jeanne’ Chen has been a fabulous friend and colleague, and her work made this dissertation possible, both in Chapter 2 and Chapter 4. Austin Hepperla provided bioinformatic support, and taught me a great deal about data analysis. Mariesa Slaughter and Aminah Wali both answered questions about protocols, and helped me work through problems in my experimental design. Jie Huang’s statistical expertise was instrumental to the analyses of Chapter 4. Dr. Dan Serber has helped with tissue culture, moving the project forward and provided critical feedback on my writing. Former members Dr. Jeremy Simon, Dr. Nick Gomez, Dr. Alexandra Arreola, and Dr. Samira Brooks all provided insight, support and friendship. All of my colleagues have been excellent scientific resources, and I consider many of them to be some of my best friends.

This work was supported by many grants. The UNC Flow Cytometry Core Facility is supported in part by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center. Due to use of the LSRFortessa, research reported in this publication was supported by the Center for AIDS Research award number 5P30AI050410. This project was supported by the V Foundation Translational Research Grant awarded to Dr. Strahl, Dr. Davis and Dr. Rathmell, R01 CA198482 awarded to Dr. Strahl, Dr. Davis and Dr. Rathmell, and K24 CA172355 awarded to Dr. Rathmell. I was supported by the UNC Medical Scientist Training Program T32 GM008719-12, NRSA Training Grant TR000085-05, and the NRSA award F30 CA192643-02. Collaborators received a UNC Lineberger Cancer Center Postdoctoral Fellowship Award (Stephen A. Shinsky, Chapter 2), NIGMS award 5T32 GM007092 (Julia V. DiFiore, Chapter 2), and R01 HL124232 (Jordan A. Shavit, Chapter 2). The confocal microscopy service in Chapter 2 was supported by UNC-CH core facility for Microscopy Services Laboratory. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: SETting the stage for cancer development - SETD2 and the consequences of lost methylation[1]

1.1 Enzymatic function and structure of SETD2

SETD2 (also known as HYPB or KMT3a) is the sole human methyltransferase that mediates trimethylation of histone H3.1 and H3 variants (Figure 1.1) [1]. Although SETD2 demonstrates biochemical evidence of mono- and dimethylation activity, in cells it seems to exclusively mediate trimethylation since SETD2 silencing results in a near complete loss of H3K36 trimethylation without decreasing mono- or dimethylation levels [1,2]. In higher eukaryotes, other enzymes, including NSD1, WHSC1 (NSD2) and SETMAR, are able to mono- and dimethylate H3K36 (reviewed in [3]). In contrast, the homolog of SETD2 in *Saccharomyces cerevisiae*, Set2, exclusively mediates all H3K36 methylation states [4]. Although extensive genetic experimentation with yeast Set2 has informed our understanding of the biochemical properties of critical SETD2 domains as well as the possible roles, the redundancy of mono- and dimethylation activities in humans offers an important caveat when extrapolating results from *S. cerevisiae*.

The functions of SETD2 have been attributed to several domains in the protein (Figure 1.2). These domains share sequence homology as well as functional similarity with those in yeast Set2. The methyltransferase activity is mediated by the conserved SET domain [4,5]. SETD2 contains two known

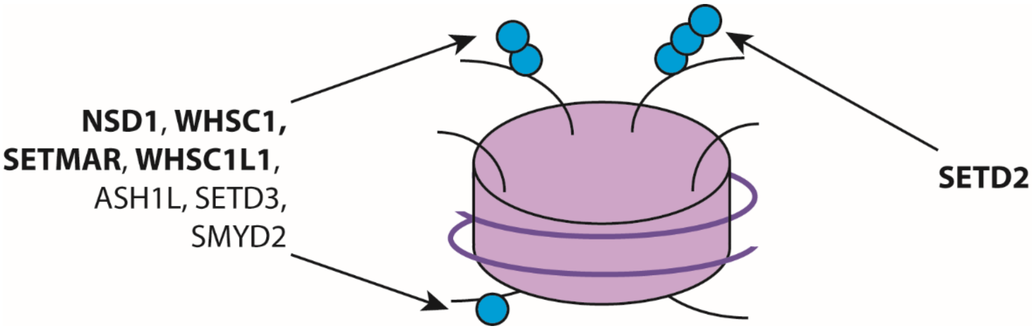


Figure 1.1: H3K36me3 methyltransferases. Methyltransferases shown to mono-, di-, and trimethylate H3K36. Those shown in bold have been shown in cell based assays and/or in vivo. Adapted from [3]

[1] This research was originally published in Cold Spring Harbor Perspectives. Fahey CC, Davis IJ. SETting the stage for cancer development: SETD2 and the consequences of lost methylation. *Cold Spring Harbor Perspectives*. 2017; 1–16. doi: 10.1101/CSHPERSPECT.a026468.

protein binding domains: SRI (Set2 Rpd1 Interacting) and WW. The SRI domain mediates association with the hyperphosphorylated C-terminal domain (CTD) of RNA polymerase II (RNAPII) [5–7]. In yeast, this interaction is required for H3K36 activity, as deletion of the RNAPII CTD decreases H3K36 methylation levels [7,8]. The WW domain, which precedes the SRI domain, may mediate intramolecular interaction [9]. Based on the property of WW domain interaction with phosphorylated proteins, this domain could also mediate other protein interactions [10]. About half of SETD2 consists of a large amino terminal domain that is not shared by yeast Set2 and is of unknown function.

1.2 SETD2 mutation in cancer

Hundreds of distinct SETD2 mutations have been identified across a wide range of human tumors, including epithelial, CNS and hematopoietic (Table 1.1, Figure 1.2) [11,12]. *SETD2* mutation was first described in clear cell renal cell carcinoma (ccRCC). In a cohort of 407 ccRCC tumors, truncating mutations were observed in twelve samples [13]. *SETD2* mutation was also found in ccRCC cell lines [14]. In The Cancer Genome Atlas (TCGA) study of ccRCC, *SETD2* was the third most commonly mutated gene with a prevalence of 15.6%. SETD2 is located on chromosome 3p, which demonstrates near universal loss of heterozygosity in ccRCC [15]. Chromosome 3p is also the location of the well-known tumor suppressor

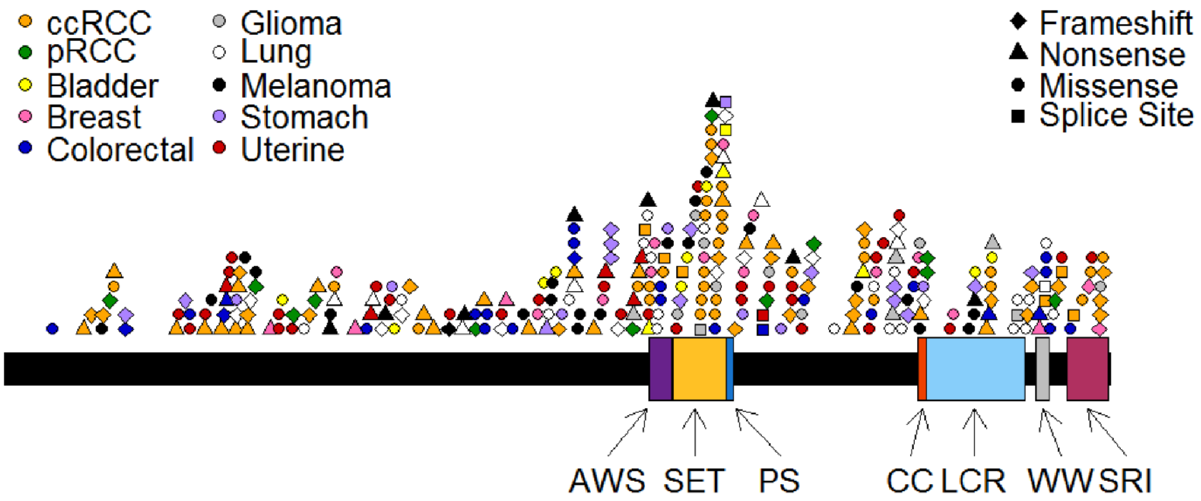


Figure 1.2: Schematic representation of SETD2 and cancer-associated mutations. SETD2 domains: AWS (Associated with SET), SET (Su(var)3-9, Enhancer-of-zeste, Trithorax), PS (Post-SET), CC (Coiled Coil), LCR (Low Charge Region), WW, SRI (Set2 Rpb1 Interacting). Mutation lists were obtained from CBioPortal on February 5th, 2016, and separated into cancer types. Duplicates were removed. Mutations are plotted by color (cancer type) and shape (mutation type). Abbreviations: Clear cell renal cell carcinoma (ccRCC) and Papillary renal cell carcinoma (pRCC).

VHL. VHL expression is lost in most cases of sporadic ccRCC, and germline mutation is associated with a high penetrance of ccRCC [16]. Mutations of *SETD2* affect the remaining allele, and frequently have a significant impact on gene function. Most mutations tend to be inactivating frameshift or nonsense mutations, although missense mutations in critical domains have been detected [17–20]. In a study of 128 sporadic ccRCC tumors that specifically examined genes known to be mutated in ccRCC, SETD2 was mutated in approximately 16% [19]. Five frameshift, ten nonsense, two splice site mutations were observed. Of three nonsynonymous missense mutations, one altered the SET domain [19]. Studies of intratumoral genetic heterogeneity also support a key role for SETD2 loss. Sequencing multiple sites in a single kidney tumor together with metastatic sites identified multiple distinct SETD2 mutations each likely to disrupt function. This convergent tumor evolution suggests that SETD2 mutation is a critical event for a subset of ccRCCs [20]. Suggestive of a link with aggressive disease, a lower level of H3K36 trimethylation was observed in metastases compared with primary tumors [21].

SETD2 mutations have also been identified in multiple other cancers. High severity *SETD2* mutation was observed in 15-28% of pediatric and 8% of adult high-grade gliomas (HGG) [22]. In contrast, *SETD2* mutations were not identified in low-grade gliomas. In addition, all tumors with *SETD2* mutation were located in the cerebral hemispheres [22]. However, mutation of *SETD2* was detected in a single

Cancer Type	Mutation % (total samples)	Reference
Clear cell renal cell carcinoma	15.6% (418 samples)	[18]
High Grade Glioma	15% (543 samples)	[22]
Uterine Carcinosarcoma Uterine Corpus Endometrioid Carcinoma	13.6% (22 samples) 9% (240 samples)	[23] [24]
Acute Lymphocytic Leukemia	12% (125 samples) 10% (94 samples)	[25] [26]
Bladder Urothelial Carcinoma	10.2% (107 samples) 6% (50 samples)	[27] [28]
Desmoplastic Melanoma Melanoma Cutaneous Melanoma	10% (20 samples) 8% (25 samples) 5.5% (91 samples)	[29] [30] [31]
Lung Adenocarcinoma	9% (230 samples) 5.5% (182 samples)	[32] [33]
Colorectal Adenocarcinoma	8.3% (72 samples) 6.1%, (212 samples)	[34] [35]
Pancreatic Adenocarcinoma	8.3% (109 samples)	[36]
Stomach Adenocarcinoma	7% (287 samples)	[37]
Papillary renal cell carcinoma	7.6% (157 samples)	[38]
Cutaneous squamous cell carcinoma	6.9% (29 samples)	[39]

Table 1.1: Cancers associated with SETD2 mutation. Cancers were selected for which the mutation rate in CBioPortal [11,12] (accessed on June 14th, 2016) exceeded 5% and a publication was available. Indicated mutation rate reflects published results. Additional cancers discussed in the text were also included.

diffuse intrinsic pontine glioma, although it co-occurred with an H3.1 K27M mutation, a common feature of these tumors [40]. Intriguingly, high grade hemispheric gliomas of children and young adults are also commonly associated with mutation of histone H3.3 (*H3F3A*) at glycine 34 [41,42]. These mutations were non-overlapping with SETD2 and were associated with reduced H3K36 methylation [22,41,43]. Gliomas also commonly harbor IDH1 mutations, which result in the generation of the 2-hydroxyglutarate (2-HG) oncometabolite [44]. These mutations are not mutually exclusive with SETD2 mutation. Although 2-HG inhibits histone demethylases, including those that can act on H3K36, it is not clear if IDH1 mutation directly affects H3K36me3 [22,45–47]. Overall, these findings indicate that dysregulation of H3K36me3 is a common event in glioma. The finding linking H3.3 mutation with reduced H3K36 methylation will be discussed in more detail below.

SETD2 mutations have also been identified in acute leukemias. In a study of early T-cell precursor acute lymphoblastic leukemia (ETP-ALL), approximately 10% demonstrated deletion or high severity mutation of *SETD2* [26]. In a separate study, SETD2 mutation was detected in approximately 6% of ALL and AML samples [48]. Mutation of SETD2 was more common in both acute lymphoid and myeloid leukemias with MLL-rearrangement compared to ALL and AML with an intact MLL gene. Further supporting *SETD2* loss as a critical event in leukemia development, *SETD2* mutations were commonly nonsense or frameshift, and approximately a quarter of samples carried biallelic *SETD2* mutations [48]. A comparison of matched primary and relapsed ALL samples suggested that mutations in epigenetic regulators as a class were more common at relapse, and this included mutations in *SETD2* [25]. In this study, SETD2 demonstrated a mutation rate of 5% in a pilot cohort and 12% in a larger validation set. The validation set contained a higher fraction of MLL-rearranged samples possibly explaining the discrepancy in mutation frequency. Suggesting that the importance of SETD2 mutation is greater in acute leukemias in children, the study of AML by TCGA identified only a single SETD2 mutation among 191 adult samples [49]. The link between SETD2 loss and MLL rearrangement is provocative since, like SETD2, MLL fusion proteins can associate with components of the transcriptional complex, and the combined alterations in these proteins may lead to transcriptional dysregulation [50,51].

SETD2 mutations have also been observed at a low frequency in a range of other tumors. 6% of melanoma and chronic lymphocytic leukemia demonstrated SETD2 alteration [30,52,53]. *SETD2* alterations were observed in high-risk, but not low-risk, gastrointestinal stromal tumors [54]. SETD2 mutation has also been described in phyllodes tumors of the breast, but not in breast fibroadenoma [55,56]. Among other genitourinary tumors, *SETD2* mutation is found in 10% of bladder tumors and the papillary subtype of renal cell carcinoma [27,38]. Although many of these mutations are monoallelic and consequently predicted to lead to haploinsufficiency, mutations are not the exclusive mechanism for modulating SETD2 activity.

Decreased H3K36me3 has also been observed in the context of non-mutant SETD2 in ccRCC [17]. miR-106b-5p, a micro RNA known to regulate SETD2, was elevated in a cohort of 40 ccRCC tumor samples; levels of this miRNA inversely correlated with SETD2 mRNA and protein levels [57]. SETD2 mRNA levels were also decreased in a subset of patients with AML and lymphoma [48].

In addition to alterations in SETD2, H3K36me3 can be lost by mutation of the methyl acceptor site in histones or by mutations in neighboring amino acids. Virtually all chondroblastomas harbor a lysine 36 to methionine variant in histone H3.3 (H3.3K36M) [58]. Expression of the H3.3K36M mutant in cells led to depletion of all H3K36 trimethylation [59]. H3.3K36M binds and directly inhibits the activity of SETD2 and the dimethylation activity of MMSET [60]. Mutations in H3.3 at the neighboring G34 residue have been described in almost all giant cell bone tumors [58] and, as previously mentioned, in high grade gliomas [40,41]. Overall, the common finding of loss or inhibition of SETD2 across a wide range of cancers suggests the importance of disrupted H3K36 methylation in cancer development.

1.3 Transcription and RNA splicing

Chromatin influences many cellular processes including transcription, replication and DNA damage repair. Many studies have linked SETD2 and transcription. In yeast, Set2 and H3K36 trimethylation are associated with gene bodies, and H3K36me3 levels correlate both with the level of transcription and the position in the gene [61,62]. H3K36me3 signals are enriched at exons, although the higher levels of nucleosome occupancy at exons may explain this difference [63]. Treatment with a transcription inhibitor causes a decrease in H3K36me3 levels [64], suggesting that active transcription is necessary for H3K36me3 placement. These data are consistent with the model that Set2 and SETD2 are targeted to elongating RNAPII.

Our understanding of the role of SETD2 in transcription is largely based on studies of Set2 in yeast. In yeast, Set2 partially functions to prevent cryptic initiation, aberrant transcription from internal sites. H3K36 methylation recruits the Rpd3C(S) complex, which includes a histone deacetylase [65,66], leading to the deacetylation of histones in gene bodies. H3K36 methylation also suppresses of the interaction between histone H3 and the histone chaperone Asf1 [67]. Preventing the incorporation of new acetylated histones maintains a hypoacetylated state, thereby stabilizing nucleosomes which decreases the chance of a segment of gene being aberrantly recognized by the transcriptional initiation complex as a promoter [65,68,69].

The role of SETD2 in transcription in higher eukaryotes is more complicated. In addition to the separation of methylation activities across multiple enzymes, genes in higher eukaryotes contain introns and are regulated by alternative splicing, and DNA itself can be methylated. H3K36 methylation levels differ based on exon utilization [70], with alternatively spliced exons having lower levels of H3K36me3 than those that are constitutively included. Altering SETD2 levels influenced the inclusion of exons in genes known to be alternatively spliced [71]. Deletion of the splice acceptor site in the β-globin intron leads to shifts in H3K36me3 signal [72], and intronless genes have lower levels of H3K36 trimethylation [64]. Chemical or RNAi mediated inhibition of splicing decreased H3K36me3 levels. Together these data suggest a close relationship between trimethylation of H3K36 and RNA splicing. Perhaps reflecting aberrant transcription or RNA processing, silencing SETD2 results in mRNA accumulation in the nucleus [73].

Several studies have explored the impact of SETD2 loss on transcription in kidney cancer. Examining primary ccRCC, H3K36me3 deficient tumors show alterations in splicing and evidence of intron retention [17]. This association was also detected in transcriptomic ccRCC data from TCGA. Similarly, differential splicing and altered exon utilization was observed in *SETD2* knockout cells [21]. However, no difference in exon usage or intron retention was observed in other studies in which SETD2 was silenced using RNAi [74]. SETD2 has also been associated with aberrant transcriptional termination [75]. In the absence of appropriate termination, RNAPII complexes can read through into neighboring genes yielding in chimeric RNAs.

Several proteins that bind H3K36 methylation offer a link between SETD and RNA splicing (Figure 1.3). LEDGF (PSIP1) exists as two isoforms, p52 and p75, based on the inclusion of six additional 3’ exons

[76]. Both forms contain a PWWP domain which interacts with di- and trimethylated H3K36. The short form, LEDGF/p52, interacts with proteins involved in alternative splicing [77]. ZMYND11, also a PWWP domain containing protein, binds H3.3K36 and associates with regulators of RNA splicing [78]. MRG15 is a chromodomain containing protein that binds di- and trimethylated H3K36 [79] and recruits polypyrimidine tract-binding protein (PTB) to alternatively spliced exons [71]. PTB then binds to silencing elements causing repression of specific exons.

In embryonic stem cells, MRG15 also recruits the lysine demethylase KDM5B to H3K36me3 marked chromatin [80]. Silencing KDM5B resulted in recruitment of unphosphorylated RNAPII to intragenic regions marked by H3K4me3, potential sites of cryptic initiation. Knockdown of KDM5B increased levels of unspliced transcripts, possibly reflecting aberrant transcription. Downregulation of SETD2 was also found to be associated with increased RNA abundance at non-initiating exons, potentially indicating transcriptional initiation from these sites [81]. Taken together, the extent to which cryptic initiation in yeast is a function of dimethylation (rather than trimethylation) or that RNA alteration in higher eukaryotic cells results from aberrant splicing (rather than cryptic initiation) are unknown.

Overall, SETD2-mediated histone modification and its interaction with specific binding partners offers an explanation for the variation in transcription and aberrant RNA detected after SETD2 loss. Although intron retention has been showed to be a mechanism of tumor-suppressor inactivation [82], how SETD2 loss facilitates tumor development remains unknown.

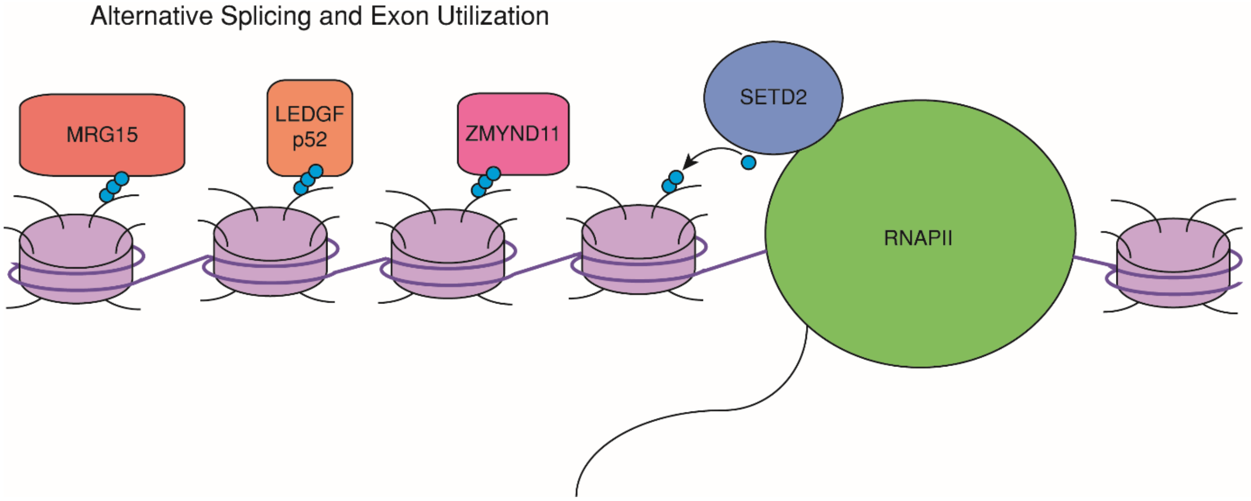


Figure 1.3: Transcription and RNA processing. SETD2 associates with RNAPII to post translationally modify nucleosomes. H3K36me3 is directly bound by readers which mediate RNA processing through downstream effectors.

1.4 Chromatin structure

Several studies have shown that H3K36me3 loss results in alterations in chromatin architecture. Silencing SETD2 impairs the recruitment of the FACT (Facilitates Chromatin Transcription) complex to transcribed chromatin which results in increased sensitivity to MNase digestion, suggesting alteration in nucleosomal interaction with DNA [81]. This effect was particularly evident at internal exons. Chromatin alterations in response to SETD2 silencing was also observed in a kidney cancer cell line [74]. A similar observation was made by examining chromatin accessibility in H3K36me3-deficient primary renal cell carcinomas using formaldehyde assisted isolation of regulatory elements (FAIRE) [17]. Overall, enhanced accessibility corresponded to regions typically marked by H3K36me3. By examining individual genic features, signal increases were most striking immediately preceding exons, suggestive of a specific effect at splice acceptor sites. Together, these studies support a link between SETD2, chromatin accessibility and splicing.

1.5 DNA replication and damage repair

Substantial evidence supports the involvement of SETD2 and H3K36 methylation in DNA damage repair by homologous recombination (HR) (Figure 1.4). SETD2 silencing resulted in a loss of HR at experimentally induced sites of double-strand breaks (DSB) [83]. In particular, SETD2 loss decreased levels of ATM phosphorylation and consequently p53 phosphorylation [84] with decreased levels of p53 transcriptional targets [84,85]. SETD2 has been shown to associate with and potentially stabilize p53, which may partially account for the difference in transcriptional targets [85]. SETD2 loss was also associated with decreased recruitment of the HR proteins 53BP1, RPA and RAD51 to chromatin [74,83,86]. This effect seems to be mediated by the histone methylation activity of SETD2 since reintroduction of a catalytically dead SETD2 mutant failed to rescue RPA or RAD51 foci formation, and depletion of H3K36me3 by overexpression of the KDM4 demethylase or H3.3K36M also delayed RAD51 foci formation [86].

LEDGF may bridge H3K36 methylation and the DNA damage response mechanism. In contrast to the role of the p52 isoform in transcription, the LEDGF/p75 isoform recruits C-terminal binding protein interacting protein (CtIP) to sites of DNA damage [87]. CtIP processes DNA ends to enable binding of RAD51 (reviewed in [88]). Depletion of SETD2 results in decreased LEDGF bound to chromatin, reduced CtIP recruitment to DSB and levels of single stranded DNA near the DSB, suggesting impaired resection

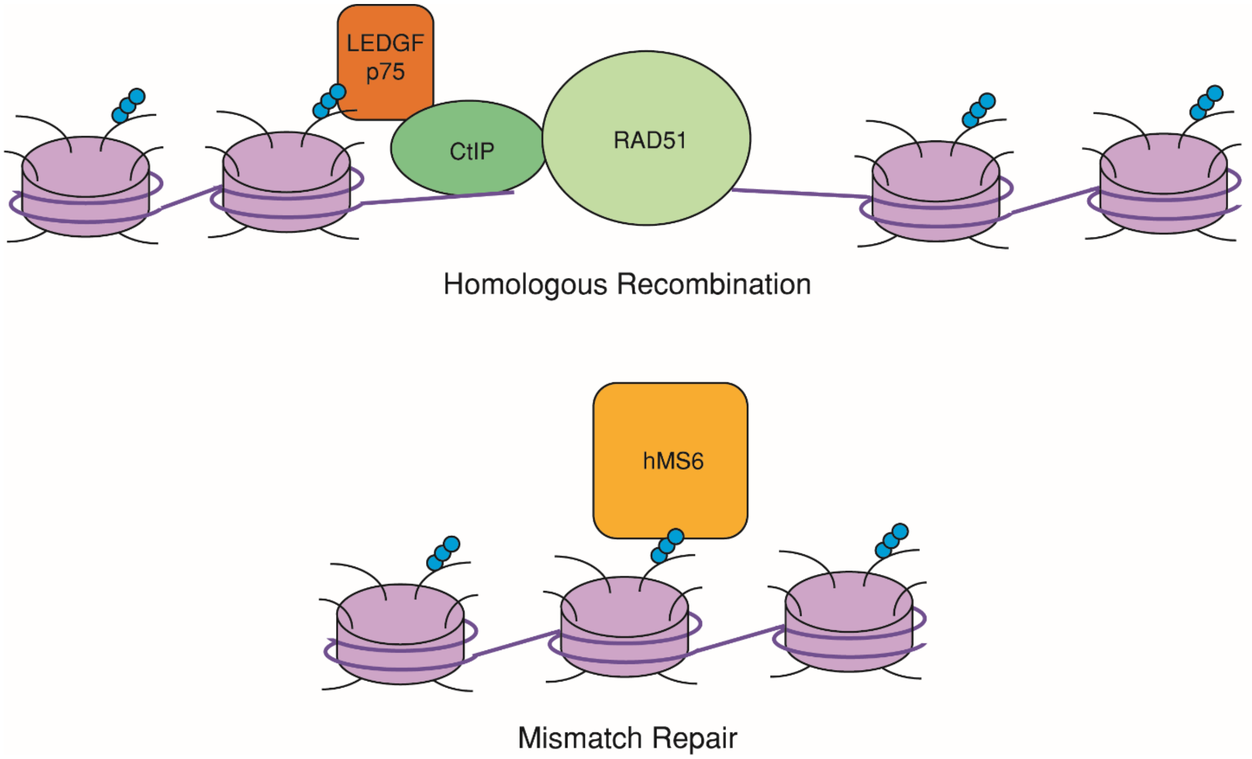


Figure 1.4: DNA damage repair. Specific H3K36me3 readers direct either homologous recombination (top) or mismatch repair (bottom).

[86]. This suggests a model in which regions marked by H3K36me3 by SETD2 are bound by LEDGF following DSB, leading to recruitment of CtIP and RAD51, and ultimately repair by HR. Whether SETD2 is recruited to sites of DNA damage that will go on to repair by homologous recombination, or conversely, that homologous recombination is more likely at regions already marked by H3K36me3 remains unclear. In contrast to H3K36me3, H3K36me2 is rapidly induced after irradiation (Fnu et al., 2011). Increased levels of early non-homologous end joining (NHEJ) factors were detected by immunoprecipitation of H3K36me2 following radiation.

In addition to HR, SETD2 is involved in mismatch repair (MMR) (Figure 1.4). hMSH6, a component of the hMutSα complex that recognizes mismatches in the genome, also contains a PWWP domain that mediates interaction with methylated H3K36 [89]. SETD2 silencing or overexpression of KDM4 decreased MSH6 foci formation associated with increased microsatellite instability (MSI) [89,90]. However, ccRCC tumors samples with biallelic SETD2 loss did not exhibit classic findings of MSI, increased breakpoints or a substantially increased mutation load, compared with tumor cells with monoallelic loss [74]. DNA breaks identified in tumors with monoalleic SETD2 loss demonstrated significantly lower levels of H3K36me3.

These data are consistent with the model that H3K36me3 marked sites are protected from breakage. Taken together, these studies suggest that H3K36 methylation functions in DNA damage repair with methylation status biasing towards different repair pathways, with dimethylation favoring NHEJ and trimethylation favoring HR and MMR. Although SETD2 deficient kidney cancers are not characterized by increased mutational level, it remains possible that intratumoral heterogeneity limits our ability to detect this feature [20].

1.6 DNA methylation and replication

Alterations in DNA methylation have been linked to SETD2 and H3K36me3 loss. The DNA methyltransferases DNMT3A and DNMT3B contain a PWWP domain enabling binding to methylated H3K36 [91,92]. DNMT3B was enriched at H3K36me3 marked gene bodies, and *SETD2* knockout reduced DNMT3B binding [92]. In this study, SETD2 loss was associated with decreased *de novo* DNA methylation, although a separate study did not observe this association [93]. Alterations in DNA methylation correlated with SETD2 loss have been demonstrated in ccRCC. In the TCGA analysis, *SETD2* mutation was associated with decreased DNA methylation at regions that are normally marked by H3K36me3 in kidney [18]. Increased chromatin accessibility was also associated with regions of DNA hypomethylation in *SETD2* mutant tumors [17]. SETD2 loss has also been associated with increased DNA methylation at intergenic regions [94]. Overall, these data suggest that H3K36me3 directs DNA methyltransferases to gene bodies but in the absence of this histone modification, methylation increases elsewhere.

H3K36me3 levels are cell cycle regulated with a peak in early S phase then declining to low levels that persist during G2/M [89]. This pattern suggests that SETD2 is most active during DNA replication. In support of a role during replication, depletion of SETD2 in kidney cancer cells slowed replication fork progression and led to an accumulation of cells in S phase [74]. However, in isogenic SETD2 knockout cell lines cell cycle differences were not observed [95].

1.7 SETD2 in cancer development and therapeutics

How SETD2 loss results in cancer development remains unknown. However, several studies have linked H3K36 methylation to aberrant differentiation or proliferation. SETD2 loss disrupts murine embryonic stem cell differentiation, possibly by altering intracellular signaling [96]. Expression of H3.3 mutants that inhibit H3K36 methylation in chondrocytes and mesenchymal progenitor cells disrupted differentiation

[60,97]. Mouse mesenchymal progenitor cells (MPCs) that stably express either wild-type or K36M mutant H3.3 formed tumors after subcutaneous injection in immunocompromised mice [97]. In renal primary epithelial tubule cells, cells considered to be the proposed progenitor for ccRCC, SETD2 knockdown resulted in continued proliferation well past the point at which these cells typically senesce [98]. In models of MLL-rearranged leukemia, SETD2 loss is associated with increased colony formation, proliferation and accelerated leukemia development after transplantation [48]. Taken together, these studies support a role of SETD2 in facilitating faithful differentiation. Interestingly, germline mutations in SETD2 as well as NSD1 (the H3K36 dimethylase, see Figure 1.1) have been associated with Sotos and Sotos-like overgrowth syndromes [99–101]. Sotos syndrome has been associated with an increased frequency of malignancy, particularly acute leukemias and lymphomas, Wilms tumor and neuroblastoma (reviewed in [102]).

The clinical implications of *SETD2* loss in cancer have primarily focused on ccRCC. *SETD2* mutation is associated with worse cancer specific survival in the TCGA dataset [103]. Additionally, *SETD2* mutation was a univariate predicator of time to recurrence, and was found at higher percentages in late stage tumors. Tumors with any of *BAP1*, *SETD2* or *KDM5C* mutation were more likely to present with advanced stage [104]. In metastatic RCC, low SETD2 expression was associated with reduced overall and progression free survival, and was an independent prognostic marker for these endpoints [105]. SETD2 expression was lower in breast tumors, compared with matched normal tissue [106,107], with expression inversely correlated with increasing tumor stage [107]. In these patients, SETD2 mRNA levels were lower in patients with poor outcomes, such metastasis, local recurrence and cancer-specific death.

Several studies have explored whether SETD2 loss sensitizes tumor cells to targeted agents. TGX221, a selective PI3Kb inhibitor, was selectively toxic to RCC cells that were mutant for both *VHL* and *SETD2*, whereas cells lacking either mutation were not sensitive [108]. Treatment with this compound resulted in decreased migration and invasion of mutant cell lines. Employing a synthetic lethality screening strategy, H3K36me3-deficient cell lines were found to be sensitive to WEE1 inhibition [95]. The proposed target of this synthetic lethal interaction is RRM2, a ribonucleotide reductase subunit. SETD2 deficiency and WEE1 inhibition each decreased RRM2 levels, and the combination resulted in further depletion. WEE1 inhibition in the context of SETD2 deficiency critically reduces the dNTP pool causing cells to accumulate in non-replicating S-phase, replication stress and cell death.

1.8 Non-histone targets

Although the focus of SETD2 research has been primarily on histone regulation, it is possible that SETD2 may have important non-histone targets. Recently, it has been reported that SETD2 targets tubulin for modification [109]. SETD2 methylates α-tubulin at lysine 40, which can also be acetylated on microtubules. This methylation occurs during mitosis and cytokinesis. When SETD2 is deleted, mitotic spindle and cytokinetic defects occur, as well as micronuclei formation and polyploidy. This suggests that SETD2 has function both in the nucleus and the cytoplasm, and may contribute to cancer development by multiple mechanisms.

1.9 Concluding Remarks and contributions of this work

Large sequencing studies have increasingly implicated mutations in epigenetic modifiers as critical events in cancer development and have identified *SETD2* loss as a key feature of multiple types of cancer. SETD2 has been implicated in many chromatin-directed nuclear processes, including transcriptional regulation, DNA damage repair, DNA methylation and replication. These effects are likely mediated by H3K36me3 binding reader proteins. Consequently, shifts in H3K36 di- and trimethylation are expected to lead to loss of appropriate reader targeting or redistribution. The relative importance of the H3K36-associated functions to cancer development remains unclear. As SETD2 mutation is associated with more aggressive cancer, it is important to fully understand the effect of SETD2 loss on oncogenesis. Vulnerabilities created by SETD2 through deregulated transcription and DNA replication may offer therapeutic strategies. The work described in this dissertation furthers our understanding of how SETD2 mutations alter activity.

Chapter 2 describes work exploring the effect of specific *SETD2* mutations on H3K36me3 and DNA damage in both yeast and human cells. First, we establish the similarities between human SETD2 and yeast Set2 by sequence and structure, and identify clear cell renal cell carcinoma mutations for further study. We then introduce these mutations into SETD2Δ human cells and Set2Δ yeast cells to characterize the effect of mutation on known functions. We show that H3K36me3 is required for DNA damage repair in human cells, while H3K36me2 is sufficient for studied yeast phenotypes.

In Chapter 3, we identify the effect of SETD2 inactivation on a genomic level, by examining changes in RNA levels through RNA-seq, chromatin organization by FAIRE-seq, and nucleosome positioning by

MNase-seq. Chapter 4 examines the effect of specific ccRCC-associated *SETD2* mutations on H3K36me3 levels using chromatin immunoprecipitation sequencing, ChIP-seq. This Chapter will demonstrate that alterations in the SRI domain of SETD2 do not alter H3K36me3 placement.

1.10 Thesis Contributions

The work described in this thesis would not have been possible without the work of many collaborators. The project described in Chapter 2 was a collaboration between the labs of Ian Davis, Kim Rathmell and Brian Strahl. The work on structure and sequence comparisons was completed by myself and Dr. Kathryn Hacker. Dr. Jordan Shavit and Andy Vo generated TALEN constructs. Dr. Hacker completed the experiments on protein stability and H3K36me3 levels. Dr. Stephen Shinsky generated data on the SET domain mutation. Julia DiFiore completed the experiments exploring protein function in yeast Set2. Dr. Yun-Chen 'Jeanne' Chiang completed the experiments on DNA damage, while I examined the effect of transcription inhibition in SETD2Δ cells. The results described in Chapter 3 were obtained in SETD2Δ cells generated by Dr. Hacker. The ChIP-seq was conducted in cells stably expressing constructs generated by Dr. Chiang. The High Throughput Sequencing Facility, J. Roach, and UNC Research Computing generated and initially processed sequencing data. All sequencing data was run through the data analysis pipeline generated by Austin Hepperla.

CHAPTER 2: Structure/Function Analysis of Recurrent Mutations in SETD2 Protein Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and Histone H3 Lys-36 Trimethylation[2]

2.1 Introduction

Cancer is increasingly characterized by alterations in chromatin-modifying enzymes [18]. *SETD2*, a non-redundant histone H3 lysine 36 (H3K36) methyltransferase [1], has been found to be mutated in a growing list of tumor types, most notably in clear cell renal cell carcinoma (ccRCC) [13,18,110], but also in high-grade gliomas [22], breast cancer [107], bladder cancer [27] and acute lymphoblastic leukemia (ALL) [25,26,48]. 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 [20]. Mutations in *SETD2* are predominantly inactivating, such as early nonsense or frameshift mutations, which lead to non-functional protein and global loss of H3K36me3 [13,17,20]. Missense mutations tend to cluster in two domains [6,13,17,18]: the SET domain, which catalyzes H3K36 trimethylation (H3K36me3) [3], and the SRI domain, which mediates the interaction between SETD2 and the hyperphosphorylated form of RNA Polymerase II (RNAPII) [6].

SETD2, and its yeast counterpart, Set2, both associate with RNAPII in a co-transcriptional manner [6,73,111]. In yeast, Set2 mediates all H3K36 methylation states (H3K36me1/me2/me3) [67] and regulates the recruitment of chromatin-remodeling enzymes (Isw1b) and a histone deacetylase (Rpd3) [68] that functions to keep gene bodies deacetylated, thereby maintaining a more compact chromatin structure [112,113] that is more resistant to inappropriate and bi-directional transcription [65,68]. The Set2/SETD2 pathway is also important for DNA repair [74,83,84,86,87,114] in both yeast and humans, as well as for proper mRNA splicing [17,71,115]. Although yeast Set2 can mediate all forms of H3K36 methylation,

[2] This research was originally published in the Journal of Biochemistry. Kathryn E. Hacker, Catherine C. Fahey, Stephen A. Shinsky, Yun-Chen J. Chiang, Julia V. DiFiore, Deepak Kumar Jha, Andy H. Vo, Jordan A. Shavit, Ian J. Davis, Brian D. Strahl, and W. Kimryn Rathmell. Structure/Function Analysis of Recurrent Mutations in SETD2 Protein Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and Histone H3 Lys-36 Trimethylation. *Journal of Biological Chemistry.* 2016; 40:21283-21295. © the American Society for Biochemistry and Molecular Biology.

SETD2 only trimethylates H3K36 [6]. Other methyltransferses (e.g., NSD2 and ASH1L) mediate mono- and dimethylation [3], indicating an increased complexity of H3K36 regulation in higher eukaryotes. 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 [91], LEDGF, which functions in DNA repair [77], and ZMYND11, which regulates co-transcriptional splicing and transcription elongation [78,116].

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.

2.2 Results

2.2.1 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 2.1A). 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 [117] whereas the SET domain of Set2 was predicted here using I-TASSER [118–121]. 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 2.1B). We then examined the conservation of amino acids previously reported to be mutated in human ccRCC [13,17,18,20] across six organisms (*H. sapiens, D. melanogaster, S. cerevisiae, M. musculus, X. tropicalis, and D. rerio*). Seven of the nine ccRCC mutations occur at residues that are conserved across all model organisms (Figure 2.1C). Additionally, three of these seven mutations occur in a region previously identified to act as the catalytic site for lysine methylation [6]. One of these mutations,

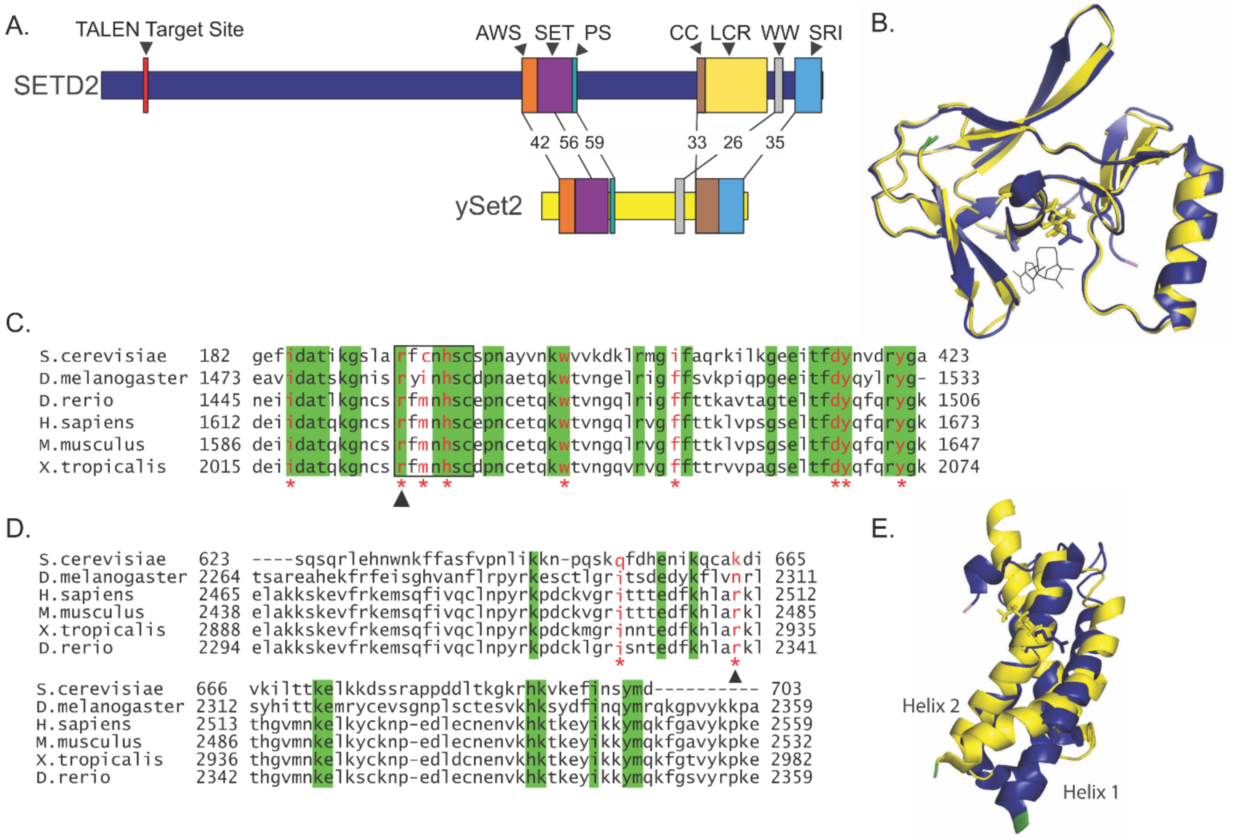

A.
TALEN Target Site
SETD2
AWS
SET
PS
CC
LCR
WW
SRI
42
56
59
33
26
35
ySet2
B.
C.
S.cerevisiae 182 gefidatikgslarfcnhscspnayvnkwvvkdklrmgifaqrkilkgeeitfdynvdryga 423
D.melanogaster 1473 eavidatskgnisryinhscdpnaetqkwtvngelrigffsvkpiqpgeeitfdyqylryg- 1533
D.rerio 1445 neiidatlkgncsrfmnhscepncetqkwtvngqlrigffttkavtagteltfdyqfqrygk 1506
H.sapiens 1612 deiidatqkgncsrfmnhscepncetqkwtvngqlrvgffttklvpsgseltfdyqfqrygk 1673
M.musculus 1586 deiidatqkgncsrfmnhscepncetqkwtvngqlrvgffttklvpsgseltfdyqfqrygk 1647
X.tropicalis 2015 deiidatqkgncsrfmnhscdpncetqkwtvngqvrvgffttrvvpagseltfdyqfqrygk 2074
D.
S.cerevisiae 623 ----sqsqrlehnwnkffasfvpnlikkn-pqskqfdhenikqcakdi 665
D.melanogaster 2264 tsareahekfrfeisghvanflrpyrkesctlgritsdedykflvnrl 2311
H.sapiens 2465 elakkskevfrkemsqfivqclnpyrkpdckvgritttedfkhlarkl 2512
M.musculus 2438 elakkskevfrkemsqfivqclnpyrkpdckvgritttedfkhlarkl 2485
X.tropicalis 2888 elakkskevfrkemsqfivqclnpyrkpdckmgrinntedfkhlarkl 2935
D.rerio 2294 elakkskevfrkemsqfivqclnpyrkpdcklgrisntedfkhlarkl 2341
S.cerevisiae 666 vkilttkelkkdssrappddltkgkrhkvkefinsymd---------- 703
D.melanogaster 2312 syhittkemrycevsgnplsctesvkhksydfinqymrqkgpvykkpa 2359
H.sapiens 2513 thgvmnkelkycknp-edlecnenvkhktkeyikkymqkfgavykpke 2559
M.musculus 2486 thgvmnkelkycknp-edlecnenvkhktkeyikkymqkfgavykpke 2532
X.tropicalis 2936 thgvmnkelkycknp-edldcnenvkhktkeyikkymqkfgtvykpke 2982
D.rerio 2342 thgvmnkelkscknp-edlecnenvkhktkeyikkymqkfgsvyrpke 2359
E.
Helix 2
Helix 1

Figure 2.1: 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.

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 2.1B). This residue is the most common site of missense mutation reported in both CBioPortal [11,12] and COSMIC [122]. The specific arginine to cysteine mutation is found in both glioma [22] and ccRCC [18]. Significantly, mutation of the corresponding residue in *S. cerevisiae* is known to affect Set2 catalytic activity [123]. 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 2.1D). However, the aligned crystal structures of yeast [124] and human [5] SRI domains display structural conservation (Figure 2.1E). 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) [5]. 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 [17,18] and is predicted to be important for SETD2-RNAPII interaction by *in vitro* peptide interaction assays [5].

2.2.2 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) (45, 46) targeting exon 3 of SETD2 were introduced into two immortalized kidney cell lines (human SV-40 immortalized proximal tubule kidney cells (HKC)(47) and 293T). Individual clones of TALEN-treated cells were isolated and loss of H3K36me3 was demonstrated by immunocytochemistry (Figure 2.2A). We verified inactivation of both alleles of SETD2 via Sanger sequencing. Representative allelic sequencing is shown (Figure 2.2B).

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 2.2C). R1625C mutant mRNA levels were also less abundant (Figure 2.2D). 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 2.2E). In contrast, the half-life of the

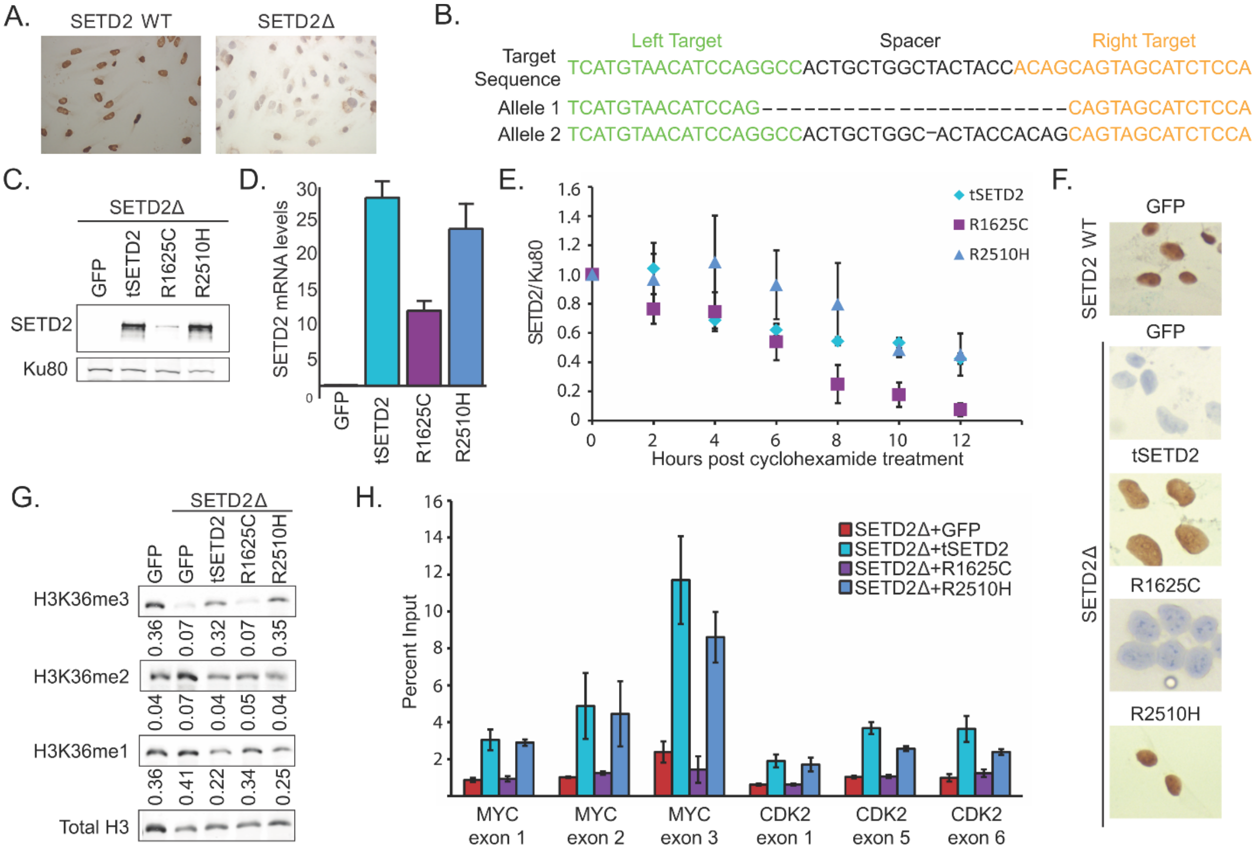


Figure 2.2: ccRCC specific mutations in SETD2 have separate effects on H3K36me3. A) Immunocytochemistry of HKC SETD2 wild-type (top) and SETD2Δ cells for H3K36me3. B) Sanger sequencing results of TALEN target sequence in exon 3 of SETD2. Two allelic variants in one HKC SETD2Δ clone are represented. C) 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 Myc (left) and CDK2 (right), displayed as ChIP signal/Input. Error bars represent standard error. Significance comparisons were made to SETD2 inactive + GFP.

R2510H mutant was unchanged (Figure 2.2E). 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.

2.2.3 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 2.2F), 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 2.2G). Monomethylation (H3K36me1) and dimethylation (H3K36me2) were unaffected by SETD2 loss or expression of SETD2 variants (Figure 2.2G). 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 [70]. Using ChIP-qPCR, we examined the H3K36me3 levels at multiple exons of two genes, CDK2 and MYC, which had previously been described [81]. As expected, SETD2Δ cells displayed low H3K36me3 levels at all sites. (Figure 2.2H). Expression of tSETD2 recapitulated the previously described pattern for H3K36me3 in wild-type cells [81] 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. 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.

2.2.4 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 2.3A). 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 2.3B).

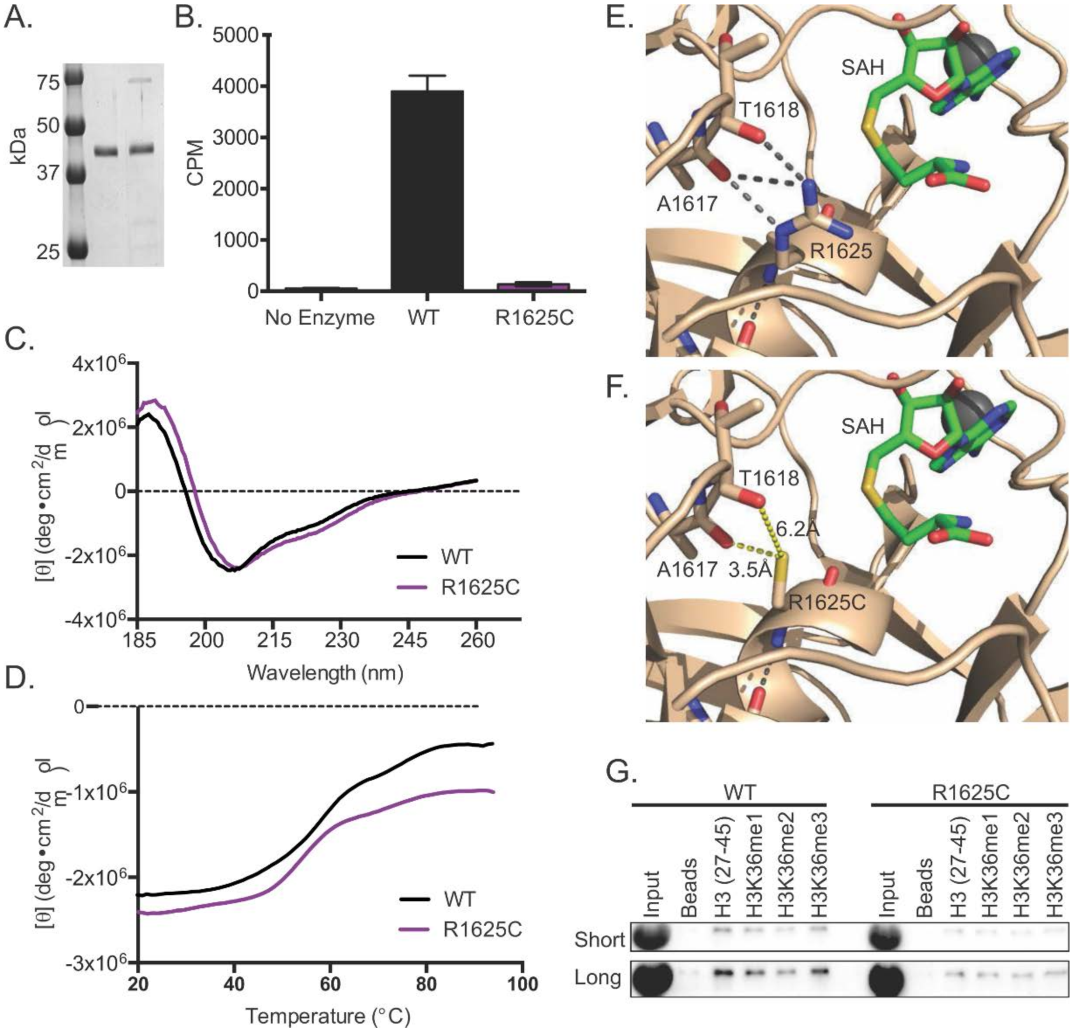


Figure 2.3: The SETD2 R1625C variant is catalytically inactive and has reduced substrate-binding capacity. A) Coomassie Blue stained SDS-PAGE gel of 1µg of purified wild type or R1625C variant SETD2 construct containing amino acids 1345-1711 (42kDa). Precision Plus Protein standards (BioRad) are annotated. 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 ^{3}H-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 unmodified, or modified as indicated. Streptavidin coated magnetic beads without peptide served as the negative control. Short (top) and Long (bottom) refer to exposure length.

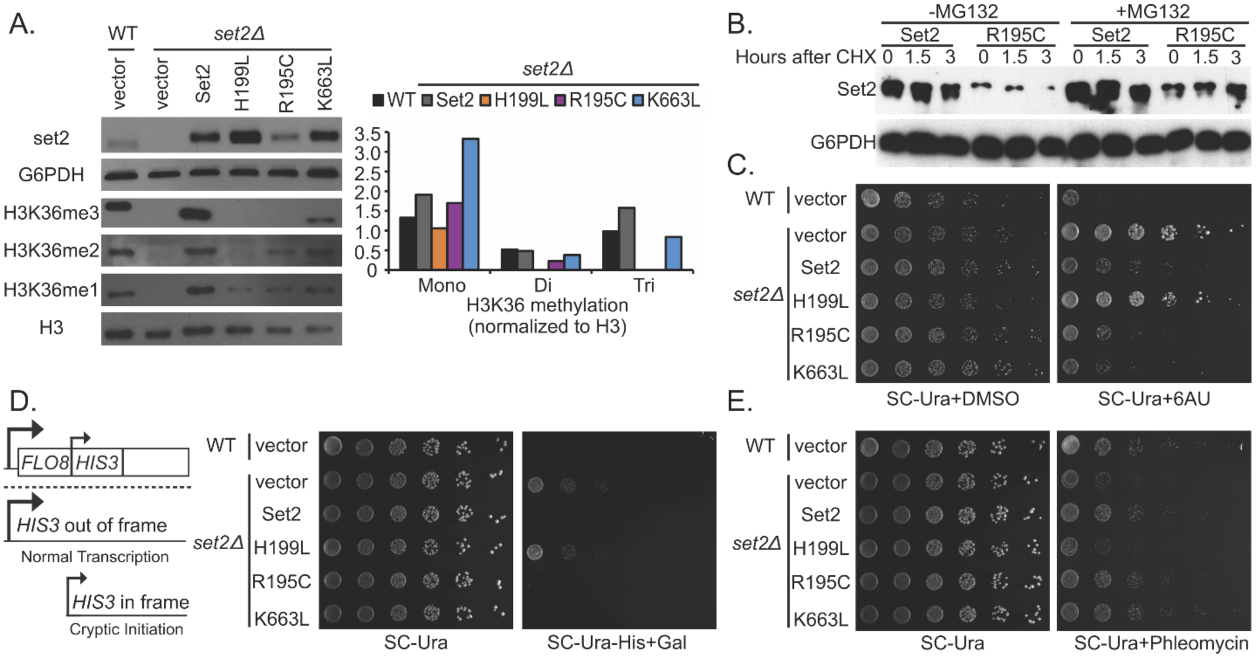
We next sought to determine why the R1625C variant is catalytically inactive. We first considered whether this mutation results in a misfolded protein, thereby inactivating the SET domain. We compared the circular dichroism (CD) spectra of the wild-type SETD2 with the R1625C variant. The CD spectra in 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 2.3C). 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 2.3D). 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 2.3E). 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 2.3E). 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 2.3F). 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 2.3G). 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.

2.2.5 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 [123], 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 [123]. 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 2.4A). 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 2.4B). 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 [123]. 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 [111]. As expected, wild-type yeast were sensitive to 6-AU (200 µg/mL), whereas set2Δ cells were resistant to this drug (Figure 2.4C) [111]. 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.



F.

Genotype	H3K36me3	H3K36me2	H3K36me1	6-AU Resistance	Cryptic Initiation	Phleomycin Sensitivity
WT	+	+	+			
set2Δ				+	+	+
set2Δ+Set2	+	+	+			
set2Δ+H199L			+	+	+	+
set2Δ+R195C		+	+			
set2Δ+K663L	+	+	+			

Figure 2.4: 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 set2∆ yeast cells expressing the indicated Set2 mutations. D) Phleomycin treatment of wild-type or set2∆ yeast cells expressing the indicated Set2 mutations. E) Cryptic initiation assay of wild-type or set2∆ yeast cells expressing the indicated Set2 mutations. F) Table summary of yeast phenotypes for each of the Set2 mutants

Cryptic initiation has been previously associated with Set2 loss [68]. 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 [68], loss of Set2 permits growth in the absence of histidine (Figure 2.4D). 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 2.4D). 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 [123,125]. 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 2.4E). 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 2.4E). 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 2.4F).

2.2.6 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 [126]. 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 2.5A).

Several recent studies have examined the effect of SETD2 loss in human cells on the response to DNA damage [74,83,84,86,87]. 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 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 SETD2∆ 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 2.5B). Quantification of these results

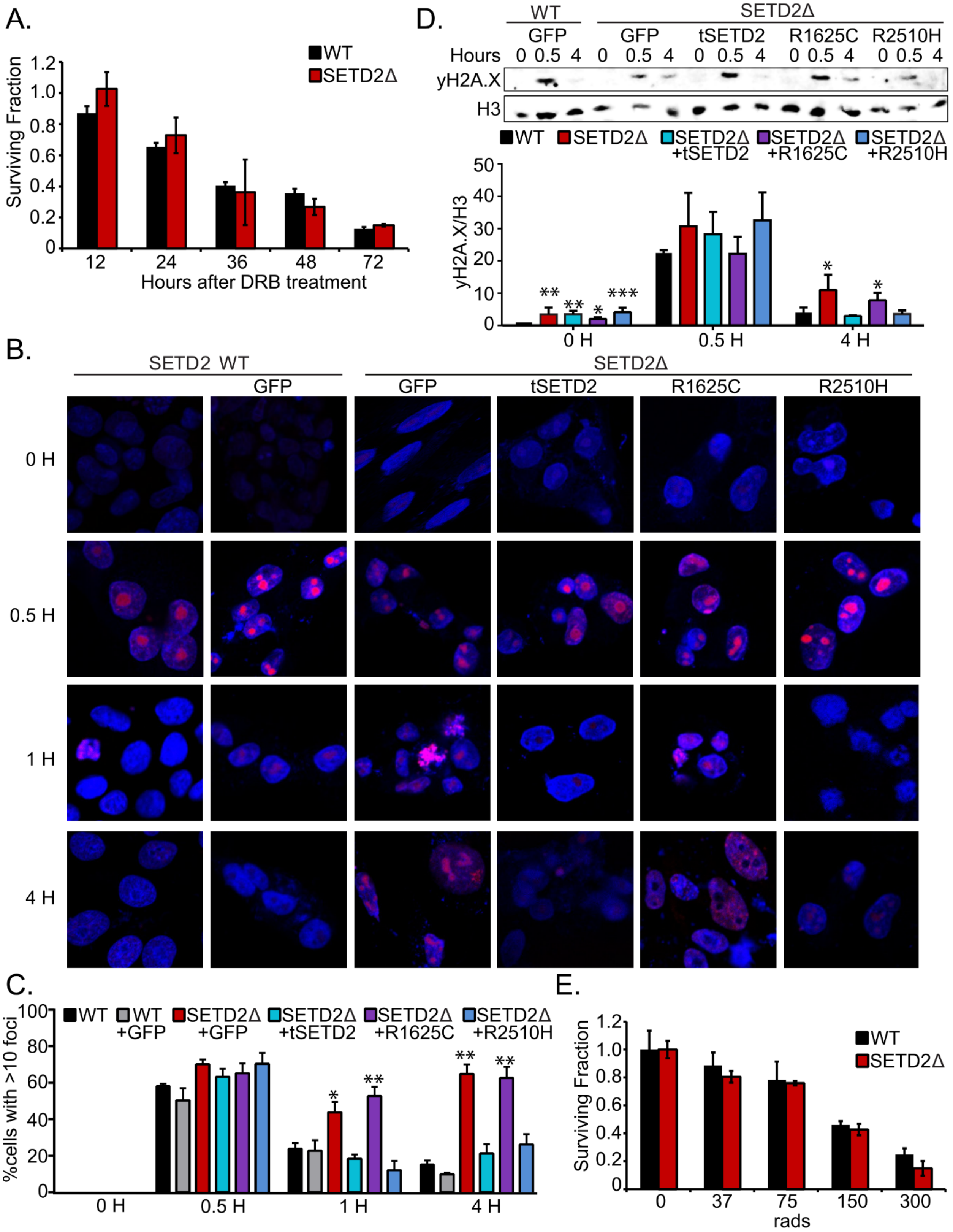


Figure 2.5: Caption on following page.

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 2.5C).

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 2.5D). 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 2.5E). 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.

Figure 2.5 H3K36me3 loss delays γH2AX 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.

2.3 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 [127] 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. 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 [112,128], 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. 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 [114]. 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 [89,90,125]. 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.

2.4 Methods

2.4.1 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

[129], and annotated with reported SETD2 mutations in ccRCC [13,17,18,20]. The structure of the SETD2 SET domain [117], SETD2 SRI domain [5] and Set2 SRI domain [124] have been previously reported. To predict the structure of the yeast SET domain, the amino acid sequence (UniProtKB, P46995) was submitted to I-TASSER using the default parameters [118–121]. The ribbon structures were aligned using the align command in the PyMOL Molecular Graphics System [130].

2.4.2 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 [131]. 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 [132]. 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 models the yeast protein in domain structure, and expression of full-length SETD2 was technically unfeasible. tSETD2 was subcloned into the pINDUCER20 vector [133]. 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).

2.4.3 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.

2.4.4 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 by the Gottschling Lab: http://labs.fhcrc.org/gottschling/Yeast%20Protocols/pprep.html.

2.4.5 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).

2.4.6 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% H2O2. 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.

2.4.7 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.

2.4.8 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). Primers used for RT-PCR are listed in table 2.1.

2.4.9 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

Target	Forward Primer (5'-3')	Reverse Primer (5'-3')
tSETD2	CACCATGACACAGGGCCA	GGGTGTCCTTGATGCTGTT
c-Myc Exon 1	GCCGCATCCACGAAACTTT	TCCTTGCTCGGGTGTTGTAAG
c-Myc Exon 2	TGCCCCTCAACGTTAGCTTC	GGCTGCACCGAGTCGTAGTC
c-Myc Exon 3	CCTGAGCAATCACCTATGAACTTG	CAAGGTTGTGAGGTTGCATTTG
CDK2 Exon 1	GTCGGGAACTCGGTGGGAG	AGAAGGCGGACCCTGGCTC
CDK2 Exon 5	CATCTGGAGCCTGGGCTGCA	TGGGGAGGAGAGGGAGGGGG
CDK2 Exon 6	CCCTATTCCCTGGAGATTCTG	CTCCGTCCATCTTCATCCAG

Table 2.1: Primers used for real-time PCR in ChIP-PCR

by growing cells in Terrific Broth II media (MP Biomedicals) at 37°C until an 0D600 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).

2.4.10 Histone Methyltransferase Assays (HMT)

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 MgCl2, 1μM ZnCl2, 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.

2.4.11 Circular Diochroism (CD) 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 A280).

2.4.12 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.

2.4.13 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 OD600 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 2.4) were serially diluted and plated on -URA-HIS plates with or without galactose for 3 days to detect growth. Growth on –URA acted as a control for equivalent growth for all the strains.

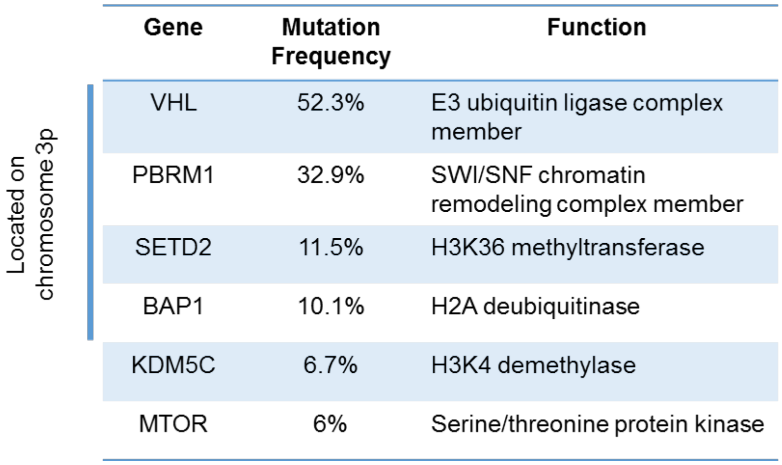
2.4.14 Immunofluorescence Staining for γH2A.X

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 4ºC 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.

CHAPTER 3: SETD2 loss results in altered gene expression and chromatin accessibility

3.1 Introduction

As described in Chapter 1, *SETD2* is a commonly mutated gene in many different cancers. Among these is renal cell carcinoma. Kidney cancer is the seventh and tenth most common form of cancer in men and women respectively in the United States [134]. It is characterized by multiple subtypes, including clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (pRCC), and chromophobe renal cell carcinoma (chRCC) [135]. Clear cell is the most common subtype, and is characterized by three major genetic alterations: loss of chromosome 3p [15], mutation of the VHL tumor suppressor [136], and mutation of chromatin modifying enzymes [18] (Figure 3.1). *SETD2* is the third most commonly mutated gene in ccRCC, and is mutated in 10-15% of tumors [13,14,17,18,20]. Many studies have attempted to understand the role of SETD2 loss in tumor development. As described in Chapter 1, SETD2 is a non-redundant histone methyltransferase in humans, and loss of this enzyme through mutation and loss of chromosome 3p results in loss of H3K36me3 in ccRCC tumors [1,17,21]. Previous work in the lab has examined H3K36me3 deficient ccRCC tumors to determine the effect of H3K36me3 loss on chromatin organization and RNA processing [17]. Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE)-seq [137] was performed on 42 ccRCC tumor samples to determine the effect of SETD2 loss on chromatin structure.



Located on chromosome 3p

Gene	Mutation Frequency	Function
VHL	52.3%	E3 ubiquitin ligase complex member
PBRM1	32.9%	SWI/SNF chromatin remodeling complex member
SETD2	11.5%	H3K36 methyltransferase
BAP1	10.1%	H2A deubiquitinase
KDM5C	6.7%	H3K4 demethylase
MTOR	6%	Serine/threonine protein kinase

Figure 3.1: Rates of mutation in clear cell renal cell carcinoma [18]. ccRCC is characterized by loss of chromosome 3p, mutation in VHL and mutation in chromatin modifiers.

Approximately 7000 500bp windows showed significantly different FAIRE signal between *SETD2*-mutant and *SETD2*-wildtype tumors, with majority of these regions showing increased signal in the *SETD2*-mutant tumors. The regions of increased accessibility overlapped regions marked by H3K36me3 in normal kidney, suggesting that loss of H3K36me3 results in increased chromatin accessibility.

In addition to changes in chromatin accessibility, H3K36me3-deficient tumors also show changes in RNA processing [17]. These tumors showed increased intron retention relative to the H3K36me3 normal tumors. This was quantified by developing an intron retention score, which compares intronic signal to exonic signal. H3K36me3-deficient tumors had drastically higher intron retention scores compared to H3K36me3 normal tumors, particularly at transcripts which also showed increased FAIRE signal. In the RNA-seq data available from The Cancer Genome Atlas (TCGA) [18], *SETD2* mutant tumors show increased levels of alternative splicing, intron retention and alternate transcription stop and start sites. Examining FAIRE signal at misspliced exons showed that the misspliced exons had increased chromatin accessibility in *SETD2* mutant tumors relative to *SETD2* wildtype tumors. Additionally, silencing of SETD2 or H3K36me3 readers has been shown to result in differential exon inclusion at individual genes [71,77] as well as alternate transcription start site utilization [81]. Recruitment of SETD2 to particular loci is transcriptionally and splicing dependent, as active splicing increases the levels of H3K36me3 at individual genes, and H3K36me3 is found more often at intron-containing compared to intron-less genes [64].

The data from TCGA and the ccRCC tumors suggested a model where loss of H3K36me3 associated with *SETD2* mutation results in increased chromatin accessibility at H3K36me3 marked regions through changes in nucleosome positioning. In this model, loss of nucleosome positioning may contribute to irregular splicing regulation, leading to altered RNA processing in *SETD2* mutant tumors. This work in tumors contributed substantially to the understanding of SETD2 in ccRCC development, but is inherently limited in its ability to draw specific conclusions for SETD2 biology. This limitation is due to the complications of tumor biology, where tumors have additional driver and passenger mutations, and have undergone selective pressure in order to survive. To study the specific effect of SETD2 loss on chromatin organization and RNA processing, we utilized the cells described in Chapter 2, which are isogenic SETD2 wildtype and SETD2Δ cell lines, and conducted MNase-seq, FAIRE-seq, and RNA-seq.

3.2 Results

3.2.1 SETD2 loss does not alter nucleosome positioning at the TSS

A nucleosome consists of a histone octomer wrapped with 147 base pairs of DNA [138]. There are two relationships between the octomer and the DNA: the translational relationship and the rotational relationship [139]. The translational relationship is the nucleosome midpoint relative to a given DNA locus, whereas the rotational relationship refers orientation of DNA helix on the surface of the histone octomer. By sequencing the DNA wrapped around the histone octomer, the translational position of the nucleosome can be determined. This DNA can be isolated using micrococcal nuclease (MNase), which cuts the DNA that is not protected by histone proteins. The DNA is then run on a gel and visualized as mono-, di-, or poly-nucleosomes. The fraction of interest is gel purified and sequenced. This sequence defines the position of the nucleosome.

As the FAIRE data from ccRCC tumors suggests a model in which *SETD2* mutation results in altered nucleosome positioning, we hypothesized that SETD2Δ cells would show widespread changes in nucleosome positioning. To test this hypothesis, SETD2 wildtype and SETD2Δ 786-O cells were treated with various concentrations of MNase. The DNA from the mononucleosome fraction was isolated from each treatment group (Figure 3.2). This was done in order to ensure equal representation of mononucleosomes released from chromatin at low levels of digestion, as well as mononucleosomes that require high levels of chromatin digestion for release. The DNA from the MNase treatment groups was then pooled, prepared into libraries and sequenced. The raw sequencing files were then processed through the MNase sequencing pipeline described in Section 3.4.6 and analyzed using the DANPOS algorithm for calling nucleosome positions and properties.

After processing the data, we first examined the nucleosome positioning at the transcription start site (TSS). This was chosen as the TSS has a characteristic nucleosome phasing that has previously been described [140]. Examining the TSS will show the quality of the MNase-seq data, and allow a comparison between SETD2 wildtype and SETD2Δ cells at this feature. In SETD2 wildtype cells, MNase signal at the TSS shows the previously described characteristic pattern of nucleosome phasing (Figure 3.3) [140].

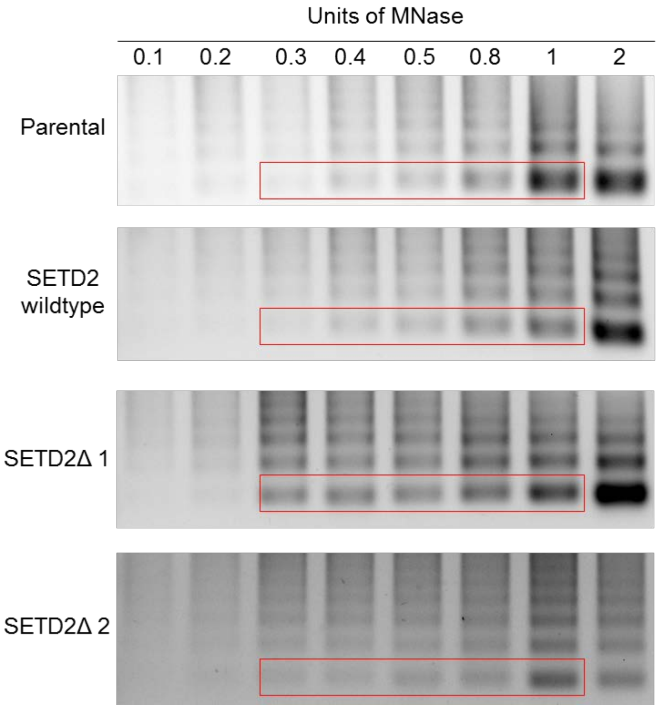


Figure 3.2: MNase treatment of 786O cell lines. SETD2 wildtype (parental and TALEN treated) and SETD2Δ cells were treated with various units of MNase. Mononucleosomes from 0.3U-1U of MNase treatment were pooled and sequenced (represented by the red box).

Centered slightly downstream of the TSS is the peak for the +1 nucleosome, which is well positioned across genes. The +2, +3, +4, and +5 nucleosomes are all easily identified, though there is loss of signal moving away from the TSS. This loss of signal is indicative of less positioning conservation across the population of nucleosomes. The -1 nucleosome is also visible, though there is an interesting phenomenon in this peak. Instead of one consistent peak, there is a second point within the -1 nucleosome. This likely represents a mixed population of nucleosomes, with some genes having a -1 nucleosome positioned ~150 bp upstream, and others with a -1 nucleosome ~200bp upstream (as estimated by eye). Overall, this result demonstrates that the MNase-seq data shows the characteristic pattern at the TSS, and is provides confidence in the quality of this data for future analyses.

We next compared the MNase-seq data at the TSS between SETD2 wildtype and SETD2Δ. Based on the widespread changes in chromatin accessibility seen in SETD2 mutant ccRCC tumors, we predicted that nucleosomes would show altered positioning in SETD2Δ cells. This was surprisingly not the case. SETD2 knockout cells showed remarkably similar nucleosome positioning at the TSS to 786O parental cells, suggesting that loss of SETD2 does not alter overall positioning at the TSS (Figure 3.4).

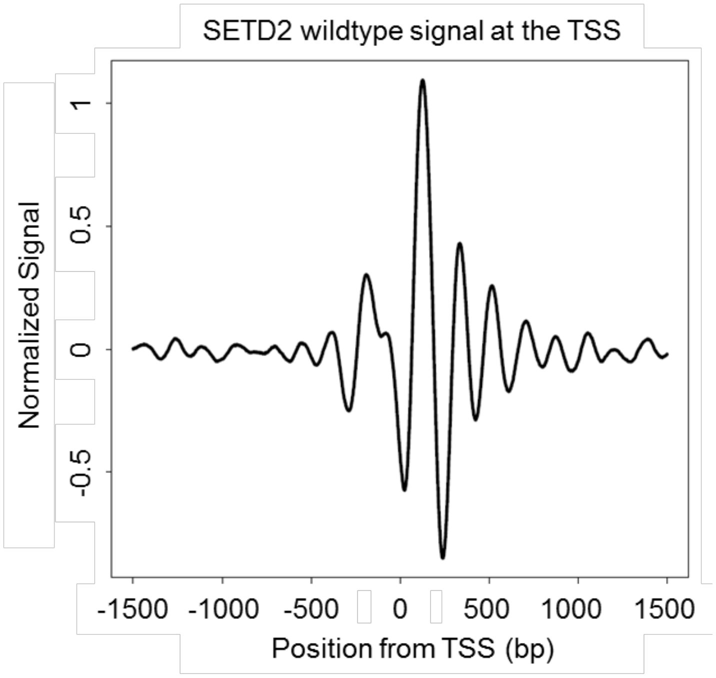


Figure 3.3: 786O Parental MNase Signal at the transcription start site. Kernal-smoothed signal for 786O parental cells centered at the transcription start site. X-axis represents bp from the TSS, y-axis represents normalized, smoothed signal.

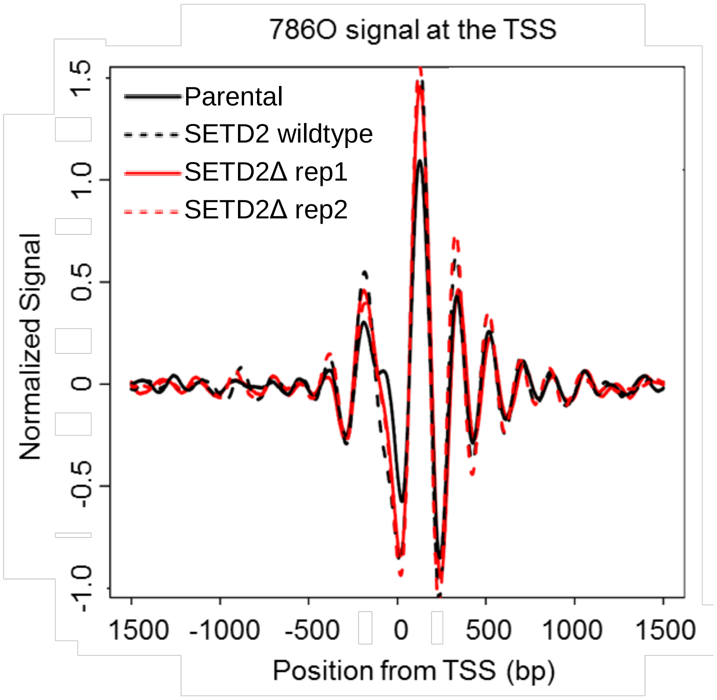


Figure 3.4: SETD2 wildtype and SETD2Δ cells show similar nucleosome positioning at the TSS. Kernal-smoothed signal for 786O cells centered at the transcription start site. X-axis represents bp from the TSS, y-axis represents normalized, smoothed signal. SETD2 wildtype samples are shown in black, SETD2Δ samples are shown in red. SETD2 wildtype refers to TALEN treated, SETD2 wildtype sample.

3.2.2 Nucleosome positioning correlates with gene expression levels

As H3K36me3 has been linked to gene expression [61,62], we next chose to examine whether the level of gene expression played a role in nucleosome positioning at the TSS. We hypothesized that the similarities seen in nucleosome positioning at the TSS were due to aggregating all genes, and that only those genes which were marked by H3K36me3 would show altered nucleosome positioning with SETD2 loss. To test this hypothesis, we utilized RNA-seq data previously generated in the lab to separate genes into quartiles based on reads per kilobase million (RPKM) values. The MNase-seq data was stratified by these quartiles to examine the effect on nucleosome positioning at the TSS.

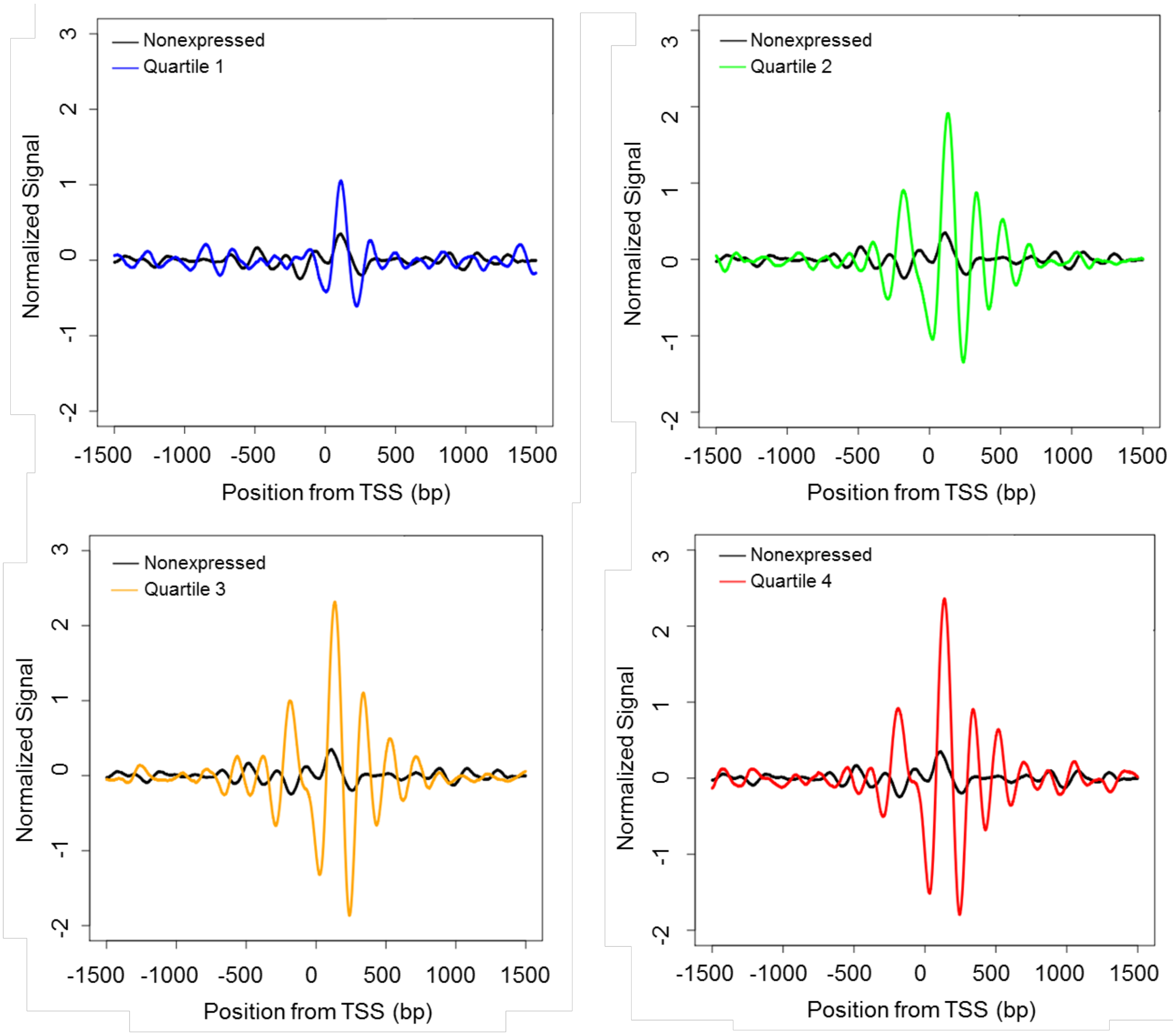


Figure 3.5: MNase-seq signal at the TSS increases with increased gene expression in 786O Parental cells. Plots of normalized kernal-smoothed MNase signal centered at the TSS, separated by gene expression quartile. All graphs show signal from non-expressed genes in black. Quartile 1 is represented in blue, quartile 2 in green, quartile 3 in yellow, and quartile 4 in red. X-axis represents bp from the TSS.

Nucleosome positioning correlates with gene expression, as the signal level increases with increased levels of gene expression (Figure 3.5). The effect of SETD2 loss on this phenomenon was examined. As in SETD2 wildtype cells, SETD2Δ cells showed increased MNase signal levels with increased gene expression (Figure 3.6). The level of signal matches that of SETD2 wildtype cells. Thus, we can conclude that loss of SETD2 does not alter signal at the TSS, regardless of the level of gene expression.

3.2.3 Internal exons are characterized by a well-placed nucleosome

SETD2 mutant ccRCC tumors showed altered chromatin accessibility at exons, particularly those with splicing defects [17]. The hypothesis was that the increase in FAIRE signal was due to loss of the H3K36me3 marked nucleosome present in SETD2 wildtype cells. To test this hypothesis, we examined MNase signal at internal exons: exons which are neither the first nor the last in the gene. These exons have a MNase peak, indicative of a well-placed nucleosome, immediately after the exon start (Figure 3.7). Surprisingly, this well-placed nucleosome is also present in SETD2Δ cells. This suggests that nucleosome positioning at internal exons is not affected by SETD2 loss.

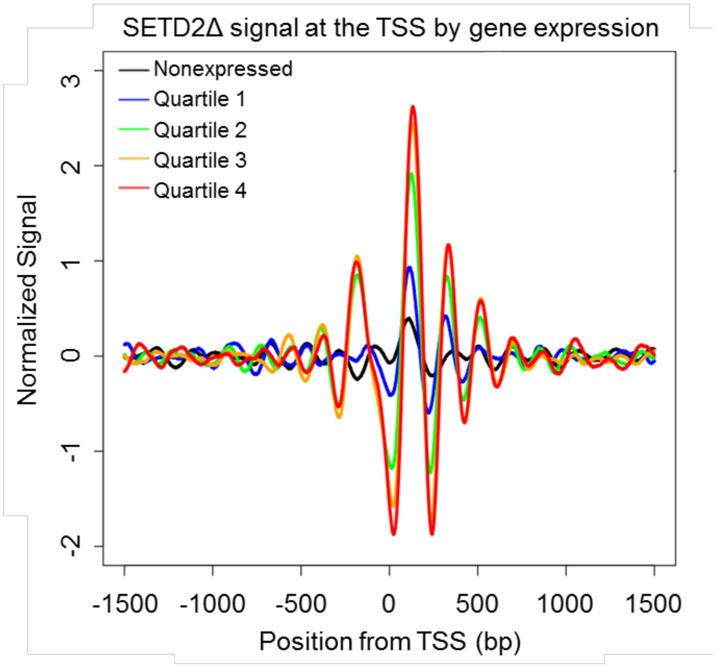


Figure 3.6: MNase-seq signal at the TSS increases with increased gene expression in SETD2Δ cells. Plots of normalized kernal-smoothed MNase-seq signal centered at the TSS, separated by gene expression quartile. Non-expressed genes are represented in black, quartile 1 in blue, quartile 2 in green, quartile 3 in yellow, and quartile 4 in red. Signal is kernal-smoothed

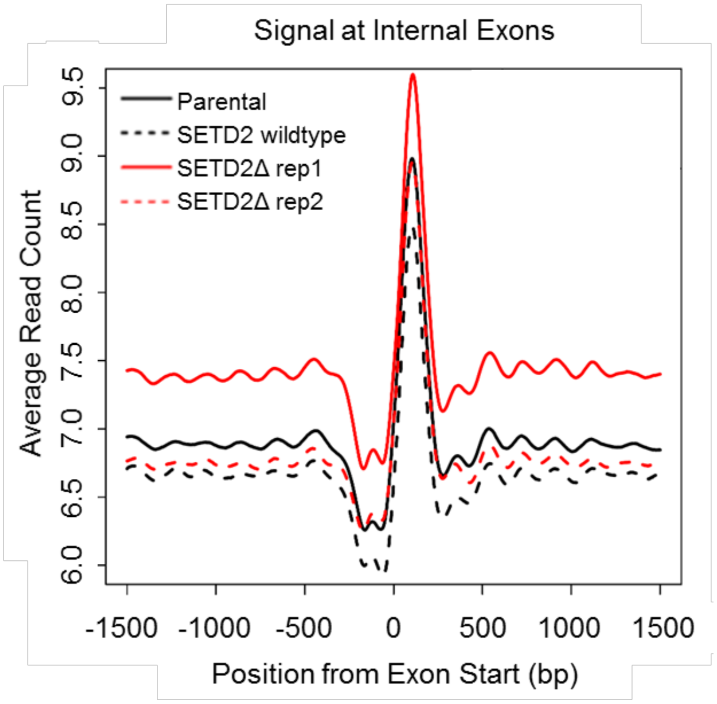
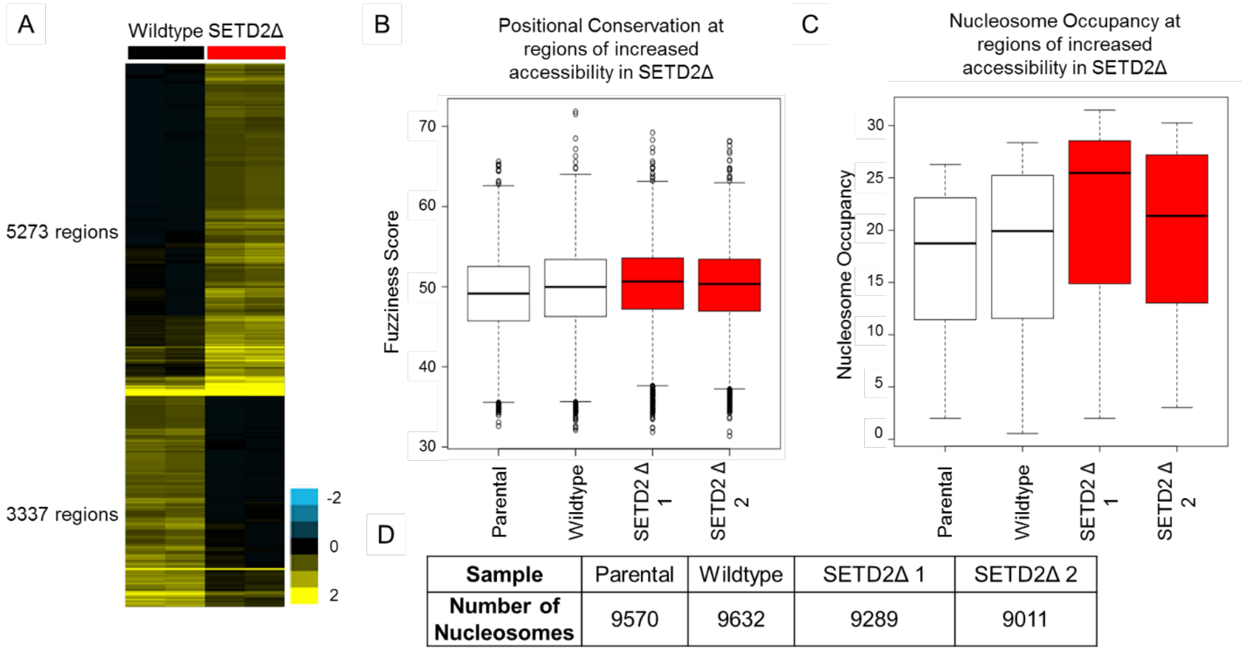


Figure 3.7: Internal exons are marked with a well-positioned nucleosome which is unaffected by SETD2 loss. Average MNase-seq signal in SETD2 wildtype (black) and SETD2Δ (red) cells is shown, centered at the start of an internal exon.

3.2.4 SETD2 loss results in changes to chromatin accessibility

As the analysis of the MNase data did not show alterations in nucleosome positioning at the examined genomic features, we decided to study chromatin using the same technique utilized in the ccRCC tumor study: FAIRE. FAIRE-seq was performed on SETD2 wildtype and SETD2Δ human kidney cells (HKC) [131]. The data was analyzed as 500bp windows. ~8500 differential regions were identified using a t-test ($p<0.05$). Of these differential regions, ~5000 show higher FAIRE signal in SETD2Δ cells, while ~3000 show higher accessibility in wildtype cells (Figure 3.8A). This suggests loss of SETD2 is associated with increased chromatin accessibility.

By combining the FAIRE data and MNase data, the properties of nucleosomes within differential FAIRE regions can be examined. Regions which showed increased accessibility in SETD2Δ cells were overlapped with nucleosomes and examined for nucleosome properties, including occupancy and positional conservation. Occupancy measures how often a nucleosome is present in a sample, and positional conservation measures the variation in the center of a nucleosome peak. Nucleosomes in regions of increased accessibility in SETD2Δ cells show decreased positional conservation (Figure 3.8B), but



Sample	Parental	Wildtype	SETD2Δ 1	SETD2Δ 2
Number of Nucleosomes	9570	9632	9289	9011

Figure 3.8: SETD2 loss results in increased chromatin accessibility with altered nucleosome properties. A) heatmap of FAIRE signal for significantly different 500bp windows (p=0.05). B) Boxplot of fuzziness score generated by DANPOS for nucleosomes in regions of increased FAIRE accessibility in SETD2Δ cells. Higher score represents decreased positional conservation. SETD2Δ samples are colored in red. C) Boxplot of Occupancy in regions of increased FAIRE accessibility in SETD2Δ cells. SETD2Δ samples are colored in red. D) Number of nucleosomes present in regions of increased FAIRE accessibility in SETD2Δ cells.

surprisingly had higher occupancy (Figure 3.8C). This suggests that nucleosomes called in the regions of increased chromatin accessibility have a higher percentage sequencing reads assigned to each nucleosome in SETD2Δ samples, but these nucleosomes are less well positioned than in the SETD2 wildtype samples. This is supported by the fact that there are fewer nucleosomes in these regions (Figure 3.8D). Overall, this suggests that in these regions, wider variation in nucleosome positioning results in more sequencing reads being assigned to an individual, less well positioned nucleosome, while the total number of nucleosomes in this region is decreased.

3.2.5 SETD2 loss results in few overall transcription changes

In addition to examining the role of SETD2 in chromatin organization, the effect of SETD2 loss on transcription was examined. As SETD2 mutation in ccRCC results in widespread transcription defects, we hypothesized that isogenic loss of SETD2 in ccRCC cell lines would result in high levels of RNA processing defects, as well as changes in gene expression. We tested this by conducting RNA-seq in both 786O SETD2 wildtype and SETD2Δ cells and HKC SETD2 wildtype and SETD2Δ cells. Using this data we are able to compare SETD2 wildtype versus SETD2Δ both within a genetic background and across different cell types.

In order to make these comparisons, we used DESeq2 [141], an industry standard. DESeq2 uses raw read counts aligned to genes as an input, estimates size factors to account for variation in sequencing depth, estimates dispersion values for each gene to account for sequencing bias across genes, before finally fitting a generalized linear model. Using this model, differentially expressed genes are determined.

The first comparison was made within the cell lines. Comparing between HKC SETD2 wildtype and HKC SETD2Δ cells identified ~360 differentially expressed genes (adjusted p<0.1) (Figure 3.9). Of these, 185 had increased expression in SETD2Δ cells, while 174 had decreased expression. When comparing the expression levels of the differential genes, it is clear that differentially expressed genes have a range of expression, and that overall expression changes are small, with only 6 genes showing greater than a 2 fold change. This suggests than in an HKC background, only a small number of genes have altered gene levels associated with SETD2 loss, and the majority of changes in gene levels are small.

Next, the same comparisons were made in 786O SETD2 wildtype and SETD2Δ cells. In the 786O background, only ~140 differentially expressed genes were identified (adjusted p-value <0.1) (Figure 3.9). Of these, 75 genes had increased expression in SETD2Δ cells, and 66 had decreased expression. Only 8

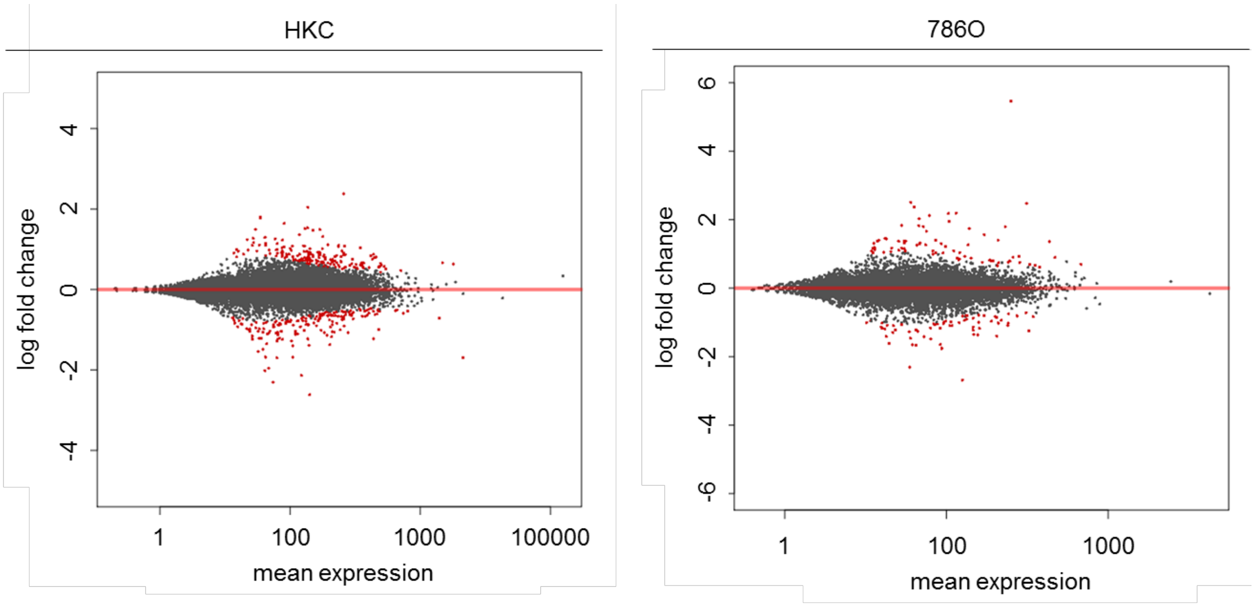


Figure 3.9: SETD2 wildtype versus SETD2Δ comparison identifies differential gene expression. MA plot showing expression changes between wildtype and knockout samples. Genes with adjusted p-value <0.1 colored in red. x-axis represents average expression of each change, y-axis represents the log fold change between SETD2 wildtype and SETD2Δ. Left: HKC, Right: 786O

of these genes showed greater than a 2 fold change. This data is very similar to the HKC data, supporting the finding that SETD2 loss does not result in widespread alterations in RNA abundance.

Comparisons were made across genetic backgrounds to test if changes in RNA abundance due to SETD2 loss were consistent between samples. The lists of differential genes found in HKC and 786O were overlapped to identify genes which are consistently altered by SETD2 loss. There was very little overlap between genes which were downregulated in SETD2Δ cells, as only 3 genes were common between 786O and HKC (Figure 3.10). Upregulated genes had similar results. In fact, 4 genes which were downregulated in 786O were upregulated in HKC, and 2 genes which were downregulated in HKC were upregulated in 786O. This result suggests that RNA abundance changes which occur with SETD2 loss are not consistent, and are affected by genetic background.

3.2.6 SETD2Δ samples have higher variation in RNA abundance than their wildtype counterparts.

As the differential genes between samples showed little overlap, similarity between replicates of SETD2Δ samples was examined. The Euclidean distance between all samples for both HKC and 786O was calculated, and the two wildtype replicates for each sample had a smaller Euclidean distance between them than the two SETD2Δ replicates (Figure 3.11A). In particular, the Euclidean distance between 786O

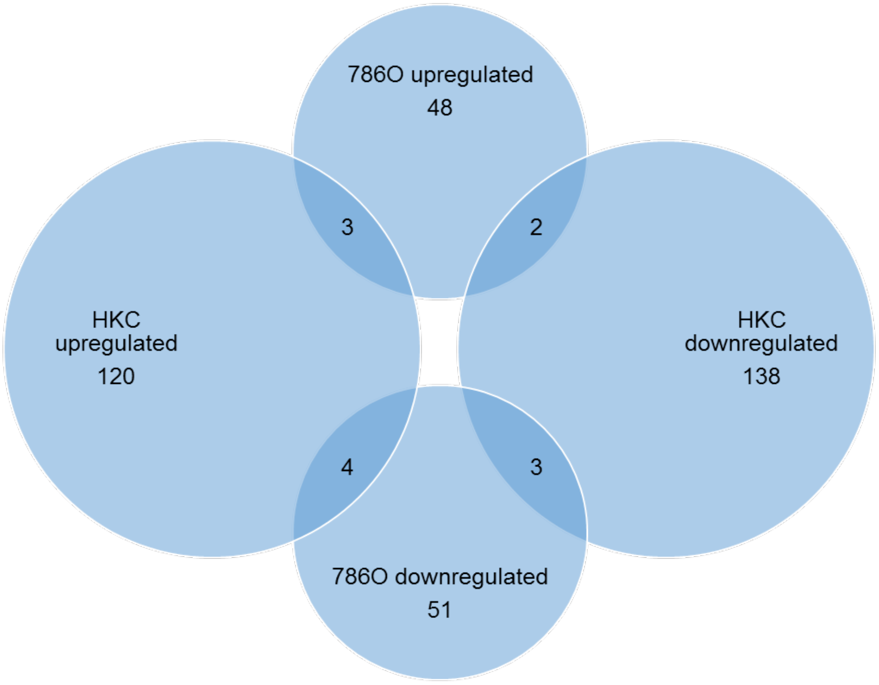


Figure 3.10 786O and HKC differential genes show little overlap. Venn diagram representing the intersection of genes identified to be upregulated or downregulated with SETD2 loss in either 786O or HKC.

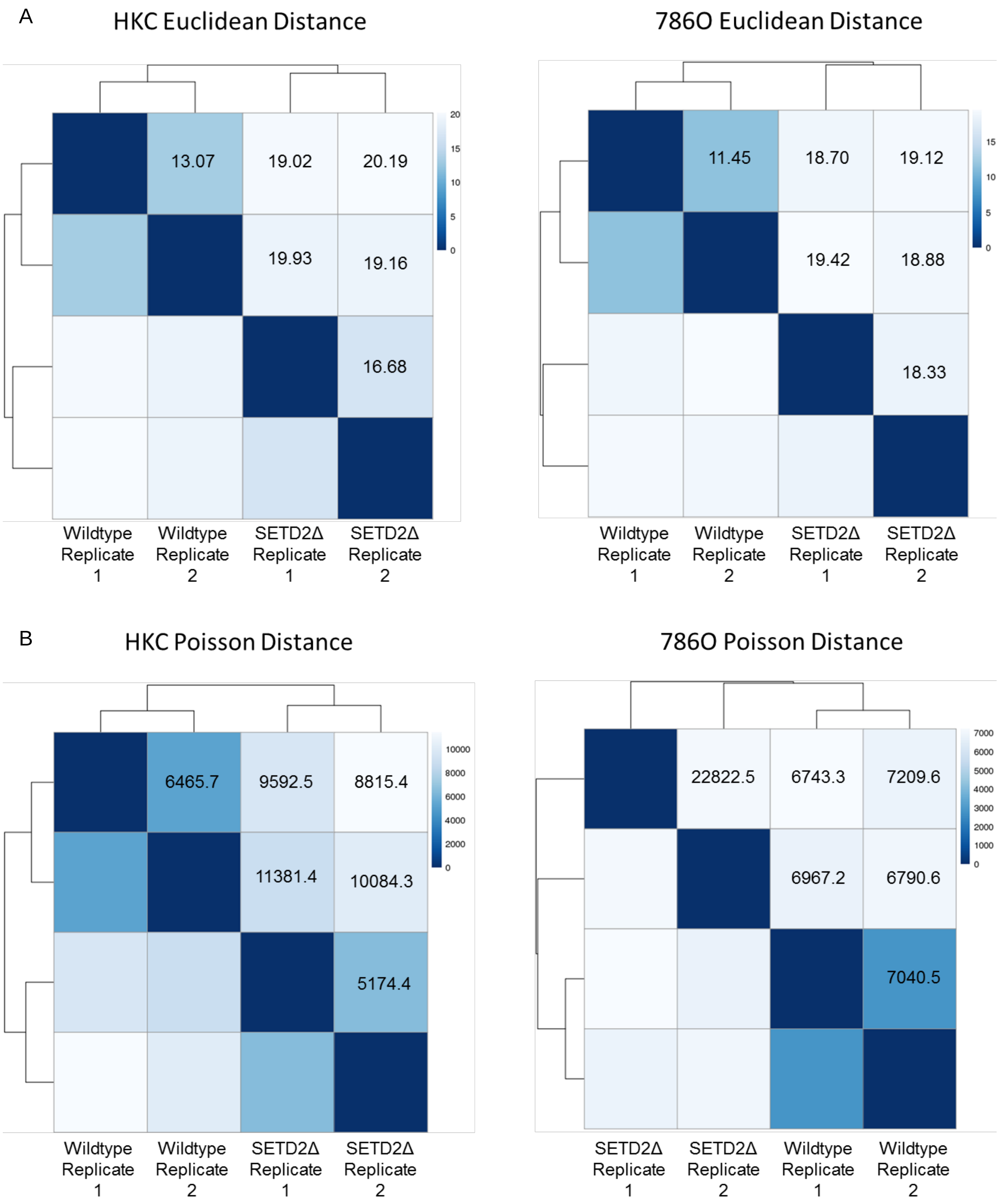


Figure 3.11: SETD2Δ replicates are more dissimilar than their wildtype counterparts. A) Euclidean distance between samples. B) Poisson distance between samples.

SETD2Δ samples was similar in magnitude to the distance between SETD2Δ and SETD2 wildtype cells. Similar analysis using the Poisson distance showed the same result (Figure 3.11B). In addition to calculating a distance metric, a principle component analysis for both cell types was completed. Principal component 1 consistently separated SETD2 wildtype from SETD2Δ samples, and accounted for 67% and 60% of the variance in HKC and 786O respectively (Figure 3.12). In both backgrounds, principle component 2 separated the two replicates for SETD2Δ, and accounted for 24% of the variance in HKC and 32% of the variance in 786O. This result indicates that though the main driver of variation between these samples is SETD2 status, a substantial driver is still variation within SETD2Δ replicates.

We hypothesized that the substantial variation within SETD2Δ replicates contributed to the low number of significant differentially expressed genes. To examine this, it is necessary to find the change in gene expression for each SETD2Δ replicate. The log2 ratio of the gene counts for each HKC SETD2Δ sample to the wildtype average was calculated, then compared across replicates. In contrast to the 6 genes which showed greater than a 2 fold change in the differential expression analysis, ~3000 genes had a >2 fold change when a single SETD2Δ sample was compared to both wildtype samples (Figure 3.13). Of these, 1336 overlap between SETD2Δ replicates. Additionally, 81 genes actually change in different directions. Unfortunately, using only one sample limits the ability to make statistically significant determinations, but this data indicates that large changes in gene expression do occur with SETD2 loss, and these changes are not consistent across samples.

This is further supported by analyses done with combined HKC and 786O datasets. Differential expression analysis shows that very few differential genes are identified in a combined data set (Figure 3.14A), and principal component analysis shows that principal component 1 separates HKC and 786O by background, and accounts for 92% of the variance (Figure 3.14B). This indicates differences in genetic background are more substantial than gene expressions changes associated with SETD2 loss.

3.2.7 Intron retention occurs in SETD2 wildtype and SETD2Δ cells

Altered RNA processing is a feature of SETD2 mutant tumors [17]. To study RNA processing, we utilized a previously published score to measure intron retention [17]. This score calculates the intronic coverage over the total gene coverage in order to determine what fraction of RNA-seq reads in a gene occur in introns. This score was calculated for each gene in the HKC and 786O samples. Surprisingly, all

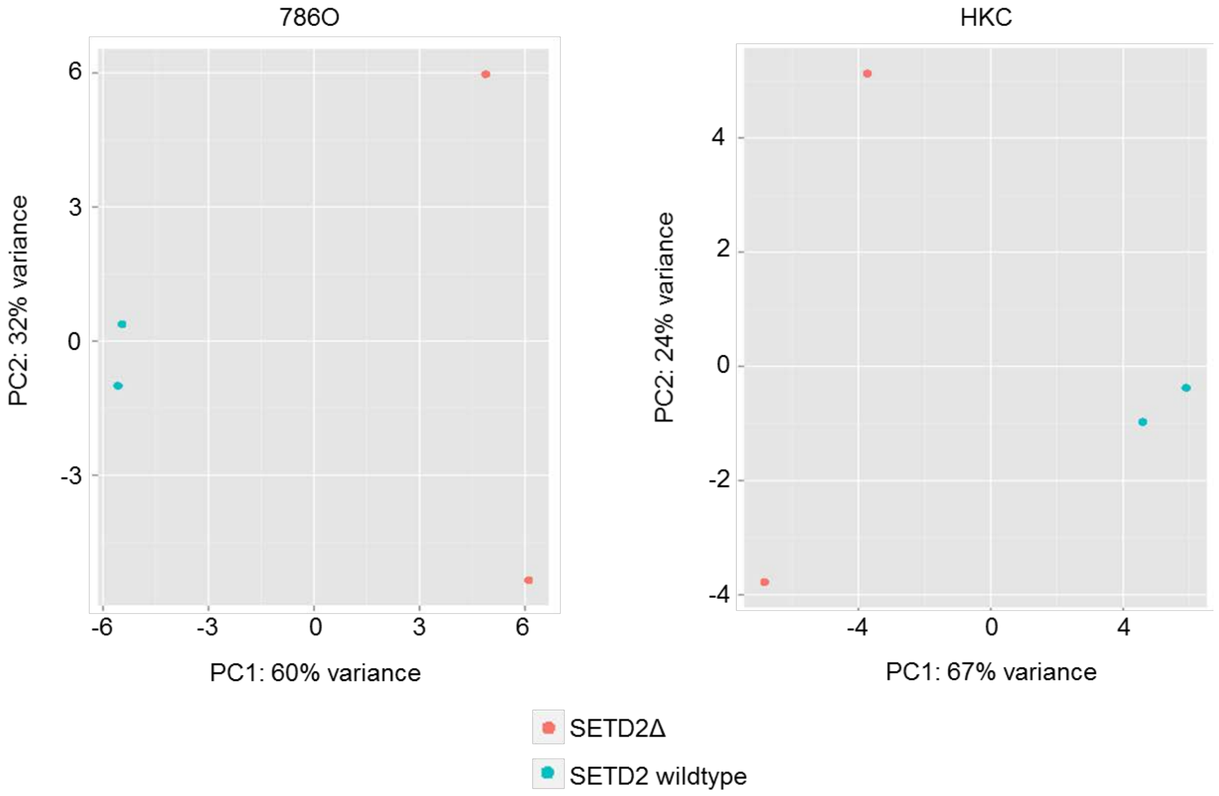


Figure 3.12: Principle component analysis of RNAseq data shows wide variation between SETD2Δ replicates. Principle component analysis of 786O (left) and HKC (right) RNAseq data.

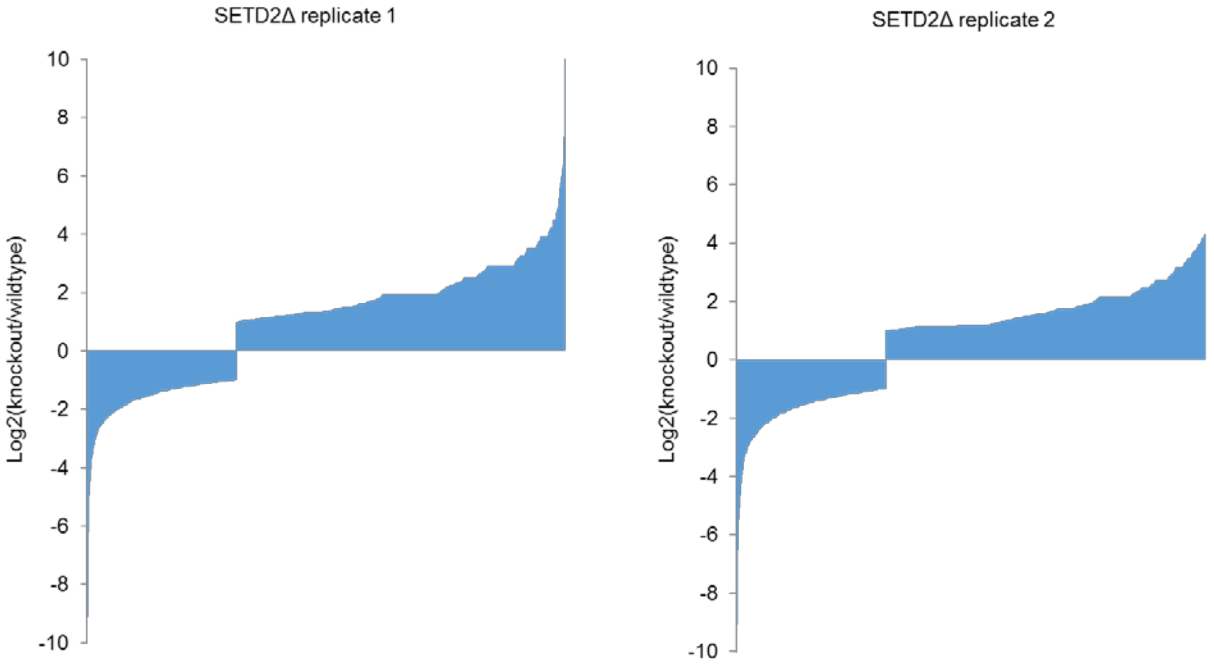


Figure 3.13: HKC SETD2Δ replicates each show high levels of gene expression changes compared to wildtype samples. Graphs show genes with >2 fold change in RNA abundance, rank ordered by expression change. Left: replicate 1, Right: replicate 2. Both replicates are compared to the average wildtype signal.

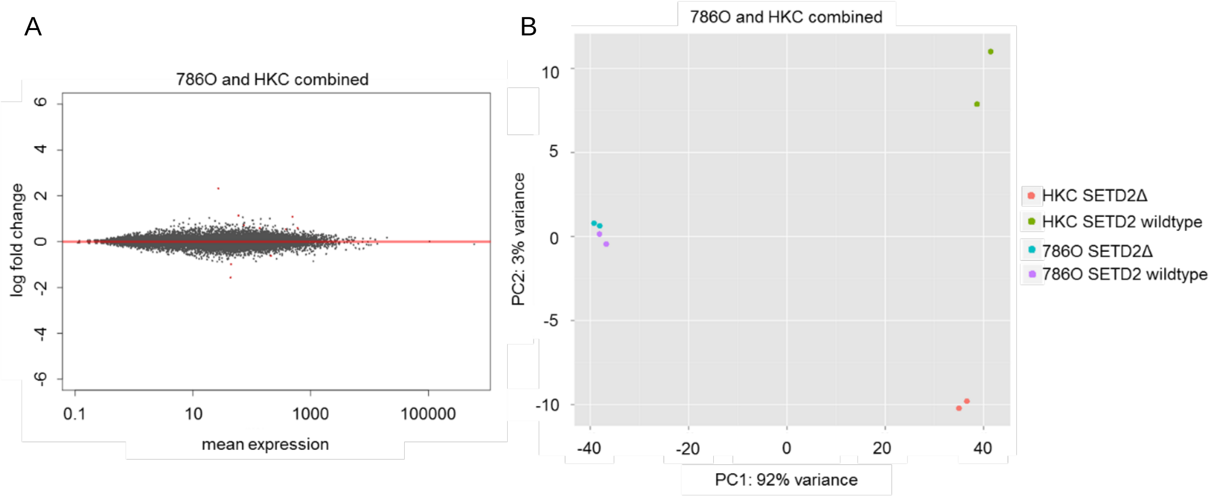


Figure 3.14: Gene expression differences between HKC and 786O are greater than those associated with SETD2 loss. A) MA plot showing expression changes between wildtype and SETD2Δ samples. Genes with adjusted p-value <0.1 colored in red. x-axis represents average expression of each gene, y-axis represents the log fold change between SETD2 wildtype and SETD2Δ samples. B) Principle component analysis of combined HKC and 786O data sets

samples showed a similar distribution of intron retention scores (IRS) (Figure 3.15).

Despite the similarities in overall distribution of IRS, comparison between 786O SETD2 wildtype and SETD2Δ cells shows 1179 genes with significantly different IRS (t-test p<0.05 post filtering). The majority of these genes show small changes in IRS (Figure 3.16A), with only 273 genes having an absolute value of difference between wildtype IRS and SETD2Δ IRS >0.05 (Figure 3.16B). Of these genes, the majority show an increased IRS in SETD2Δ cells. When comparing changes in intron retention score to gene expression levels (as measured by reads per kilobase million (RPKM)), it is clear that genes which show high changes in IRS between wildtype and knockout are also lowly expressed in both SETD2 wildtype (Figure 3.16C) and SETD2Δ (Figure 3.16D) samples. This data suggests that changes in intron retention do occur with SETD2 loss, however, the majority of genes are not substantially affected.

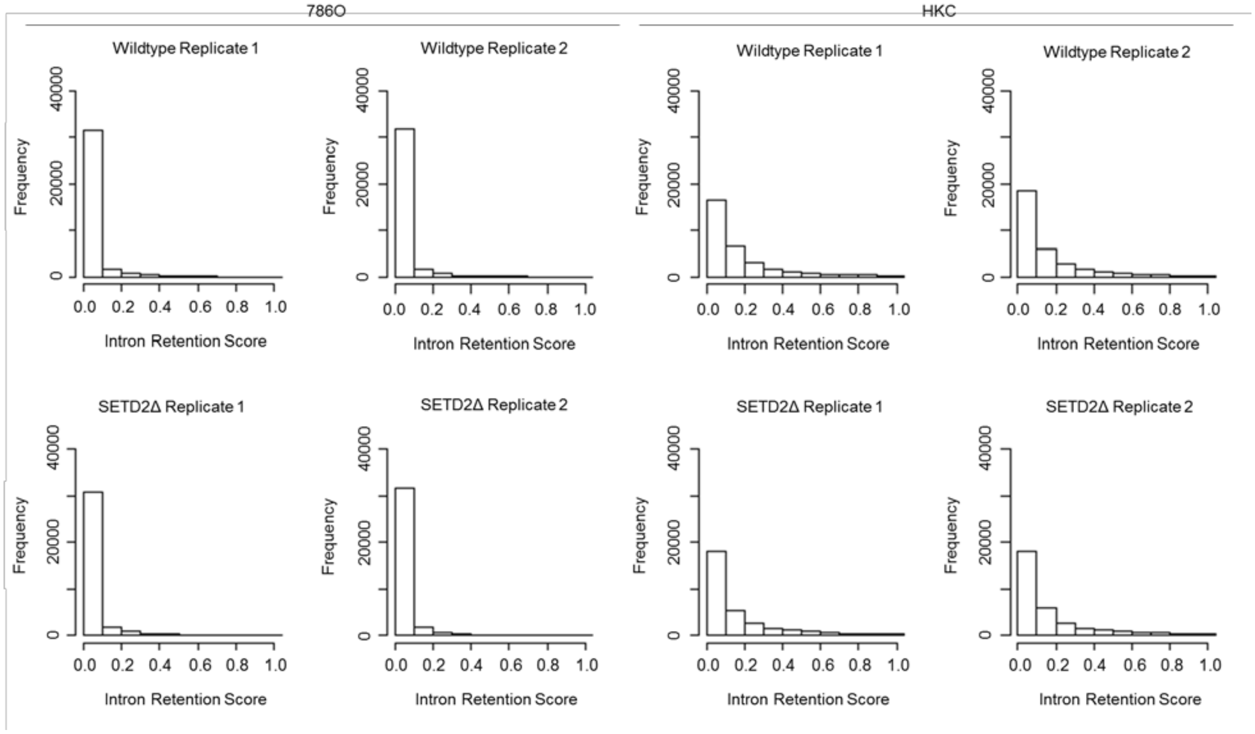


Figure 3.15: Intron retention score is similar across SETD2 wildtype and SETD2Δ samples. Left: Histogram of 786O IRS for each sample. Right: Histogram of HKC IRS for each sample.

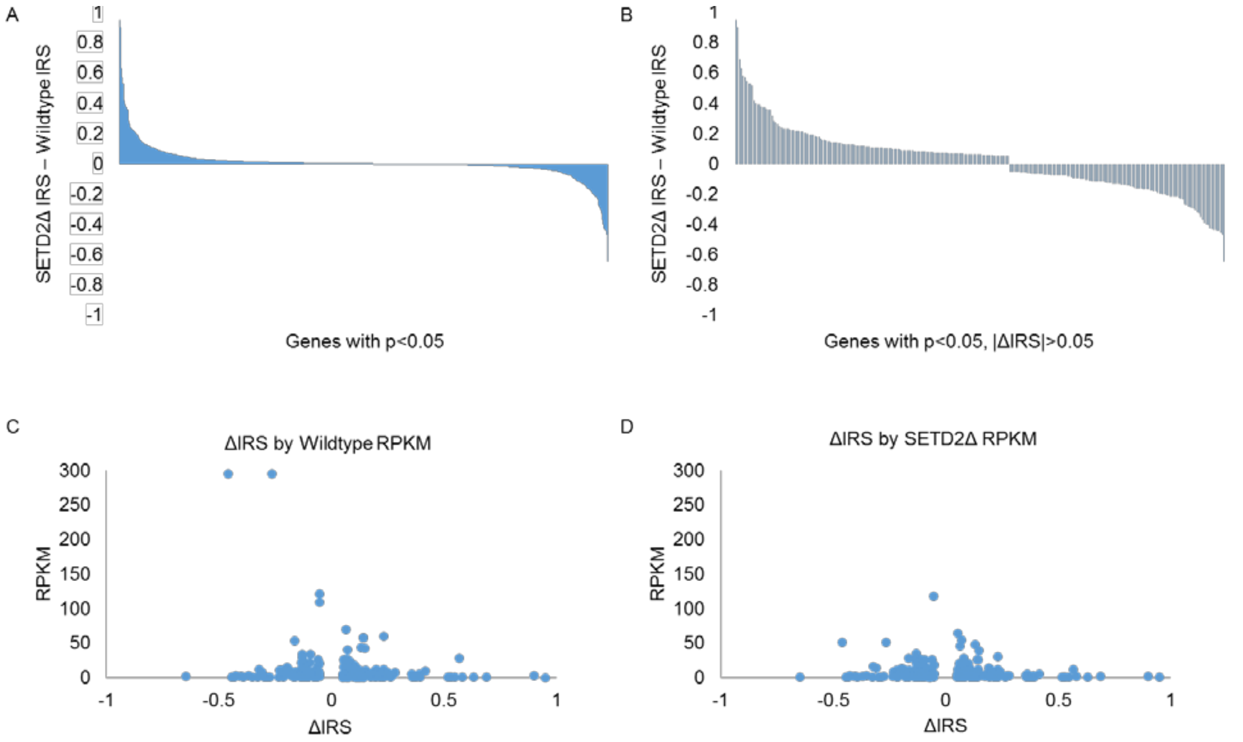


Figure 3.16: Intron retention differences between SETD2 wildtype and SETD2Δ samples are small and occur at lowly expressed genes. A) Rank-ordered representation of knockout IRS – wildtype IRS (ΔIRS), where each line represents 1 gene. B) Filtered results of A for those with a |ΔIRS|>0.05 C) Genes from B plotted by ΔIRS versus SETD2 wildtype average RPKM D) Genes from B plotted by ΔIRS versus SETD2Δ average RPKM

3.3 Discussion

Multiple lines of evidence indicate the importance of *SETD2* mutation in ccRCC development. *SETD2* mutation correlates with worse cancer-specific survival in ccRCC [103]. Studies which explored intratumor heterogeneity have identified independent mutations in *SETD2* across multiple subsections of an individual tumor, suggesting that *SETD2* mutation is a critical and selected event in ccRCC development [20]. Understanding how SETD2 loss contributes to ccRCC development is key to understanding this disease, as well as identifying new therapeutic opportunities.

Previous work has shown that in ccRCC tumors, H3K36me3 loss results in altered chromatin accessibility and RNA processing [17]. The work in this chapter builds on this, and suggests that the role of SETD2 in these processes may be somewhat more nuanced than previously suspected. The initial hypothesis for changes in chromatin accessibility seen in ccRCC tumors was that H3K36me3 marked nucleosomes were no longer properly positioned, particularly at internal exons which were misspliced. This model is not supported by the MNase-seq results. As internal exons and TSS show no alterations in nucleosome positioning associated with SETD2, these loci do not rely on H3K36e3 marking for positioning cues. Changes in chromatin accessibility are still found in SETD2Δ cells, indicating that this is a feature common to SETD2 loss, not just a feature of ccRCC SETD2 mutant tumors. Furthermore, nucleosome properties at regions of increased chromatin accessibility are altered. Studies in other models have shown than changes in chromatin accessibility can occur without changes in nucleosome positioning due to altered nucleosome properties [142]. In that study, regions with increased FAIRE signal had MNase signal suggesting well positioned nucleosomes. It is possible that in ccRCC tumors, the increased chromatin accessibility is due to altered nucleosome properties, rather than changes in positioning. This model would be supported by the chromatin accessibility and nucleosome positioning data described above.

H3K36me3 deficient ccRCC tumors also showed widespread alterations in RNA processing, with nearly 25% of expressed genes showing alterations [17]. Many studies conducted since this result was published have examined the role of SETD2 in RNA processing, with conflicting results. In one study of SETD2Δ cells, altered exon utilization and differential splicing occurred [21]. However, in a study of SETD2 silencing, exon usage and intron retention were unaltered [74]. We have found that intron retention is increased with SETD2 loss, but the number of genes as well as the level of expression of these genes is low. Additionally, large changes in RNA abundance did not occur. One potential explanation for this is cell cycle related. Certain genes have periodic expression which correlate with cell cycle stage [143]. H3K36me3 is enriched over the gene body of these genes. H3K36me3 levels are also highest in G1, and decrease with the cell cycle [89]. As the RNA for this study was taken from an unsynchronized population, it is possible that cell cycle changes in H3K36me3 and associated changes in gene expression are lost in the total RNA. Future studies may consider synchronizing cells to ensure comparisons are made within the same cell cycle stage.

Our studies also suggest substantial variation between SETD2Δ replicates. As described in Chapter 2, the role of H3K36me3 in DNA damage is an area of active study [84,86,90]. As the cells used for this study were single cell sorted, grown as clones, then passaged for experiments, DNA damage may have occurred and been improperly repaired during this process. This would result in clonal variation between replicates. As shown above, changes in genetic background contribute more to variation in RNA abundance than loss of SETD2 (Figure 3.14). To further explore this phenomenon, studies using an inducible SETD2 loss through protein degradation are being explored. Using a system in which SETD2 loss occurs in a systematic, time course manor would allow comparisons to be made minutes, hours, and days after SETD2 loss. This would allow the dynamics of H3K36me3, RNA processing and chromatin accessibility changes to be further explored, as well as provide context for early ccRCC tumor development.

The results described in this chapter suggest that SETD2 has a role in chromatin accessibility and RNA processing, but also paint a complicated picture of regulation. Further analysis of all generated datasets are necessary for a complete understanding of the effect of SETD2 loss on these features. Identifying misspliced exons and other alterations in RNA processing will further our understanding of the role of SETD2 in RNA processing. Additionally, once genes with aberrant processing are identified, FAIRE and MNase signal can be examined at these exon or genes. This will allow us to expand the model of nucleosome positioning and chromatin accessibility in RNA processing originally proposed to more accurately describe biological events in the cell.

3.4 Methods

3.4.1 Cell Lines Used

786O and HKC SETD2Δ generation was previously described in Section 2.4.2. For MNase-seq, 786O parental, TALEN-treated, and two SETD2Δ cell lines were used. TALEN-treated refers to cells which were treated with both TALEN vectors, but did not show loss of H3K36me3 by dot blot and were verified to have intact SETD2 sequence by allelic sequencing. For FAIRE-seq, HKC parental, HKC left-TALEN treated, HKC SETD2Δ 1, and HKC SETD2Δ 2 cell lines were used. HKC left-TALEN treated refers to cells which were treated with only the left targeting TALEN construct, and did not show loss of H3K36me3 by dot blot and were verified to have intact SETD2 sequence by allelic sequencing. RNA-seq was done using all the above cell lines.

3.4.2 MNase Treatment

786O parental, 786O TALEN-treated, 786O SETD2Δ replicate 1 and 786O SETD2Δ replicate 2 cells were grown on 15cm plates, isolated and washed with cold PBS, and resuspended in resuspension buffer (RSB: 10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$). The suspension was incubated on ice for 10 minutes, before adding NP-40 (final concentration 1%) and nutating on ice 30 minutes. The samples were centrifuged (3000RPM 4°C 5 minutes), washed 2x with RSB, and resuspended in 200uL of MNase reaction buffer (10mM Tris-HCl pH7.5, 5mM $MgCl_2$, 5mM $CaCl_2$ 0.1mM PMSF, 0.5mM DTT). The OD 260 for each sample was determined using the Nanodrop 1000 Spectrophotometer, and samples were diluted to a OD 260 of 1. The desired units of MNase were added (0.1, 0.2, 0.3 0.4, 0.5, 0.8, 1, 2 units), and samples were placed at 37°C for 10 minutes. The digestion reaction was stopped by addition of 10mM EDTA/EGTA, then RNAse and Proteinase K treated. The DNA was extracted using a phenol-chloroform extraction, then run on a 2% agarose gel to assess digestion levels. After digestion was verified, DNA was run on a new agarose gel, and mononucleosome DNA from 0.3U – 1U treatment groups was pooled, prepared into libraries (described below) and sequenced.

3.4.3 FAIRE

HKC parental, HKC left-TALEN treated, HKC SETD2Δ replicate 1, and HKC SETD2Δ replicate 2 cells were treated with 1% formaldehyde for 7 minutes, inactivated with 125mM glycine, scraped to collect, washed with cold PBS, and resuspended in FAIRE lysis buffer (10mM Tris-HCl pH 8.0, 2% Triton X-100, 1% SDS, 100mM NaCl, 1mM EDTA). The sample was sonicated to an average fragment size of ~400bp. DNA was phenol-chloroform extracted, and prepared into libraries as described below.

3.4.4 cDNA generation for RNA sequencing

Total RNA was isolated from HKC parental, HKC left-talen treated, HKC SETD2Δ replicate 1, and HKC SETD2Δ replicate 2 using TRIzol reagent (Ambion RNA 15596-026) following the manufacturers protocol. RNA was Ribo-Minus treated (Ribo-Minus Eukaryote System v2, Life Technologies A15026), fragmented (Ambion RNA by Life Technologies Fragmentation Reagents, AM8740), first strand synthesized using random primers, RNase H treated and second strand synthesized. 100ng of cDNA was used for library preparation as described below. 786O sequencing data was previously generated by Dr. Kathryn Hacker [144].

3.4.5 Library preparation

Sequencing libraries for MNase DNA, cDNA, and FAIRE DNA all followed the standard Illumina sequencing protocol. For MNase-seq, 100ng of DNA was used as input. For RNA-seq, 80-100ng of cDNA was used as input. For FAIRE-seq 100ng of DNA was used as input. DNA was blunted, purified using Ampure XP beads, A-tailed, adapter ligated, purified using Ampure beads two additional times, before PCR amplification (12 cycles for MNase, 12 cycles for RNA, 15 cycles for FAIRE) and one additional purification. Final libraries were submitted as a pool to the UNC High Throughput Sequencing Facility (HTSF) for sequencing. MNase-seq utilized 4 lanes of paired-end 50bp reads. RNAseq utilized 2 lanes of paired-end 50 base pair reads. FAIRE-seq utilized 1 lane of paired-end 50 base pair reads.

3.4.6 Sequencing pipeline and data processing

After fastq files were obtained from the HTSF, the files were processed to remove adapters using TagDust [145]. Paired-end sequences were synchronized to each other, then aligned to the genome using Bowtie [146]. The bowtie output files were convered to BAM files using samtools [147], then filtered for properly mapped reads. These files were used to generate downstream files for additional analyses. For MNase-seq, BAM files were used by DANPOS [148] (version 2.1.3) to call nucleosomes and nucleosome properties. Differential FAIRE regions were called by comparing signal in 500 base pair genomic windows between SETD2 wildtype and SETD2Δ samples using a t-test, and a p-value cutoff of 0.05. Nucleosomes within a FAIRE region required a 1 base pair overlap between the 500 base pair window and the DANPOS called nucleosome.

3.4.7 MNase signal at specific genomic features

MNase signal plots at specific genomic features were generated using R. Wiggle files were input into Zinba [149] to create .coord files for the desired feature using the coord.spbc function. This .coord file was collapsed into a single profile for the feature, which was graphed using standard R parameters. For kernel-smoothed graphs, the ksmooth function within R was utilized. Features files were created by downloading genomic information from the UCSC table browser [150], and formatting for Zinba use. TSS files were generated from TSS data downloaded from UCSC, with 3 kilobases added to each end of the annotated TSS, accounting for standedness. Internal exons were generated by downloading an exon list, removing the first and last exon for each gene, and add 1.5 kilobases to each end of the exon start, accounting for strandedness.

3.4.8 RNA-seq analysis

Differential expression analysis was determined using DESeq2 [141] with the default parameters. HTSeq gene counts were used as the input for DESeq2 [151]. Differential genes were called at adjusted p-value<0.1. Euclidean distance, Poisson Distance and Principal Component analysis were all completed using DESeq2 with rlog transformed data. Reads per kilobase-million (RPKM) for individual genes were calculated using an in-house script. Intron retention scores were calculated as previously described [17].

CHAPTER 4: SRI alteration does not affect genomic placement of H3K36me3

4.1 Introduction

The overall goal of this dissertation is to contribute to the understanding of the role of SETD2 mutation in ccRCC development. Chapter 2 described the generation of isogenic SETD2 wildtype and SETD2Δ cells, as well as the reintroduction of tSETD2 and two point mutations. Chapter 3 studied the effect of SETD2 loss on chromatin organization and RNA processing. These studies have contributed to our understanding of the effect of SETD2 loss on specific phenotypes, but did not explore the effect of loss and mutation on the original described role of SETD2 as a histone methyltransferase.

SETD2 mutations in ccRCC can be placed into 3 general categories: early inactivating, SET domain mutations, and SRI alterations (Figure 4.1A). In Chapter 2, we examined the effect of TALEN inactivation (representative of early inactivating), one SET domain mutation (R1625C), and one SRI domain mutation (R2501H) on methyltransferase activity. The R1625C mutation disrupts the catalytic activity of SETD2 (Figure 2.2, [152]). The R2510H mutation in the SRI domain did not impair global restoration of H3K36me3, as measured by immunohistochemistry and western blot (Figure 2.2, [152]). This level of analysis does not, however, determine how genomic placement of this mark is affected.

In addition to the SRI point mutant, a new mutation was generated: T2457*. This mutation converts T2457, the first amino acid of the SRI domain, to a stop codon, effectively removing this domain. We hypothesized that alteration of the SRI domain, though mutation or deletion, would alter genomic positioning of the H3K36me3 mark. To test this hypothesis, we conducted chromatin immunoprecipitation for H3K36me3 across our cell lines.

4.2 Results

4.2.1 The SRI domain of SETD2 is required for interaction with RNAPII but dispensable for methylation

As described in Chapter 2, we have generated a series of SETD2Δ cells using TAL effector nucleases [152]. Using these cells, we reintroduced a truncated SETD2 construct, tSETD2 (amino acids

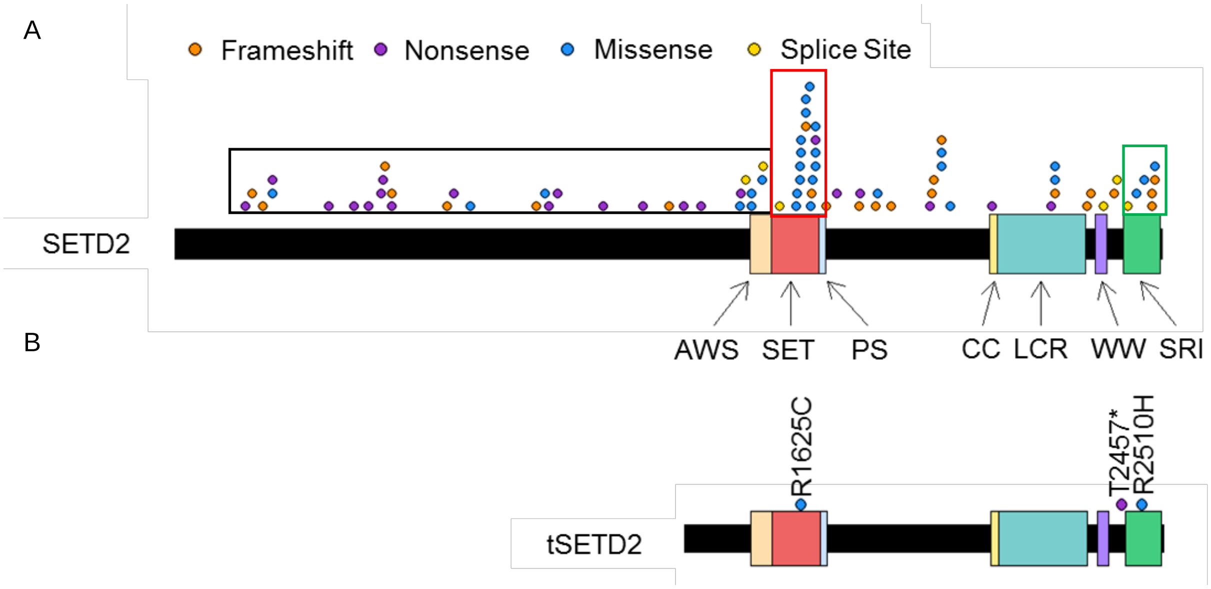
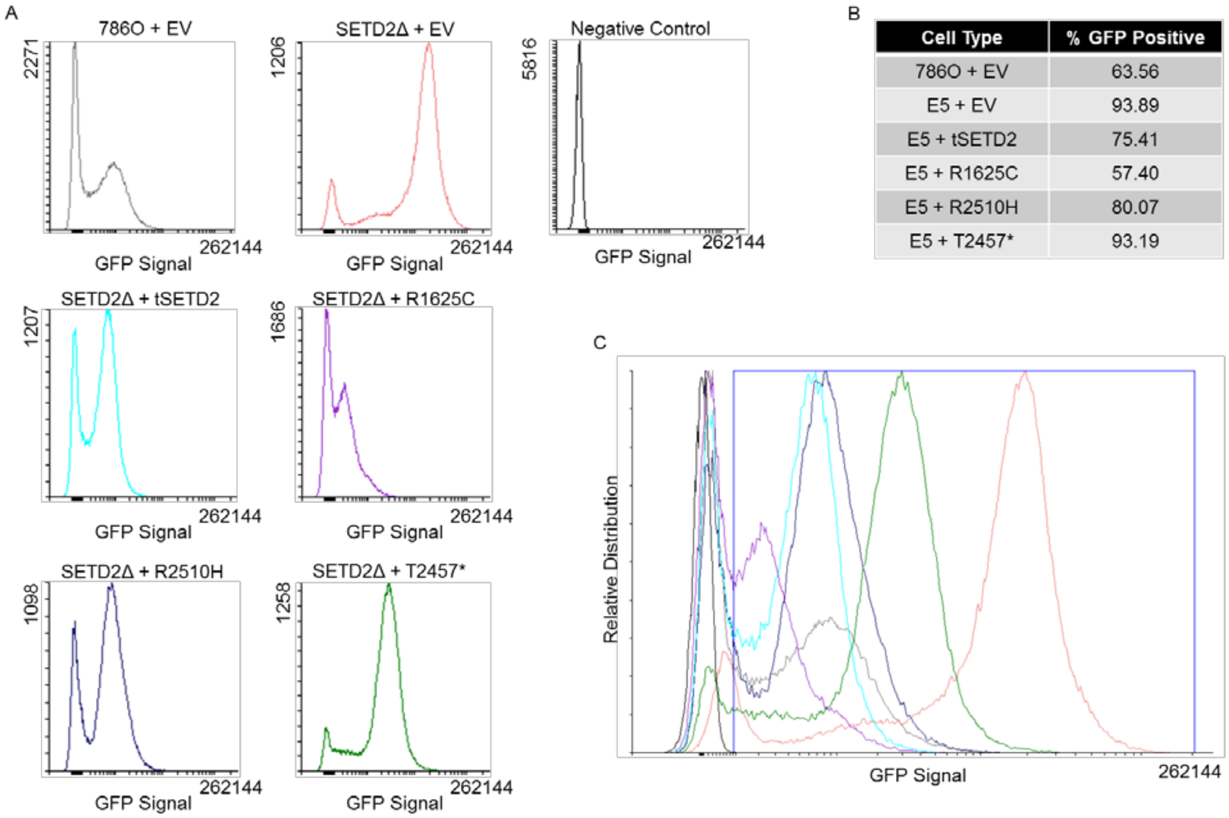


Figure 4.1: SETD2 mutations in ccRCC fall into three categories. A) ccRCC SETD2 mutations annotated in cBioPortal [11,12]. Colored boxes represent mutation types. Black = early inactivating, Red = SET domain, Green = SRI domain B) Schematic of tSETD2 construct and mutations used for ChIP-seq

1323–2564). tSETD2 includes all known functional domains of SETD2 (Figure 4.1B) and transient expression of tSETD2 restores H3K36me3 globally, [152]. In contrast to the experiments of Chapter 2, which were conducted in cells transiently expressing tSETD2 and mutants, we transduced SETD2Δ cells with lentivirus and selected for stable expression of tSETD2 and the mutant constructs. The stable expression construct has an HA-tag on the N-terminus of tSETD2, and a C-terminal GFP sequence. GFP and tSETD2 (or mutant tSETD2) are co-transcribed, but separated by a 2A peptide. The 2A peptide is an 18 amino acid sequence found in the aphthovirus foot-and-mouth disease virus [153]. This sequence interacts with the ribosome to promote hydrolysis of the peptidyl(2A)-tRNA^{Gly} ester linkage, releasing the synthesized polypeptide, while proceeding to translate the remaining RNA sequence [154]. This system has been developed in many cellular backgrounds as a way to ensure equal expression of multiple peptides [155,156].

Using the tSETD2-2A-GFP system, we transduced 786O SETD2Δ cells with the construct of choice, selected with antibiotics, and used flow cytometry cell sorting to select GFP positive cells. GFP expression levels post-sorting were re-examined by flow cytometry (Figure 4.2). The percent GFP positive cells varied between samples. Based on this, the level of GFP expression was verified by western blot (Figure 4.3). GFP expression levels tracked with the percent GFP positive cells determine by flow cytometry, with SETD2Δ+EV having the highest level of GFP, and SETD2Δ+R1625C having the lowest.



Cell Type	% GFP Positive
786O + EV	63.56
E5 + EV	93.89
E5 + tSETD2	75.41
E5 + R1625C	57.40
E5 + R2510H	80.07
E5 + T2457*	93.19

Figure 4.2: Stable Expression of SETD2 mutants with GFP shows substantial GFP positive population. A) Histogram of GFP Signal in individual samples and a non-transfected negative control. B) Table of percent of the total population which is GFP positive C) Overlapped tracks of A, with gating box used to determine percent positive shown in blue.

Given that GFP levels should match tSETD2 levels, we also blotted for the HA tag present in the constructs. Surprisingly, tSETD2 had the lowest levels of expression, despite similar levels of GFP to other samples (Figure 4.3). The level of HA tag was normalized to the tubulin loading control, and comparisons between samples show large levels of variation. SETD2Δ+tSETD2 has the lowest expression, while SETD2Δ+T2457* expression is very high.

We next examined the restoration of H3K36me3 in SETD2Δ cells by stable expression of the various mutants using western blotting (Figure 4.4). As previously found with transient expression, SETD2Δ and cells stably expressing the R1625C point mutation both lack H3K36me3 (Figure 2.2, [152]). Additionally, stable tSETD2 expression restores H3K36me3 to wildtype levels, as does the stable expression of the R2510H point mutation. We predicted that the SRI deletion would not be capable of trimethylation. Surprisingly, stable expression of the T2457* mutant in SETD2Δ cells restores H3K36me3.

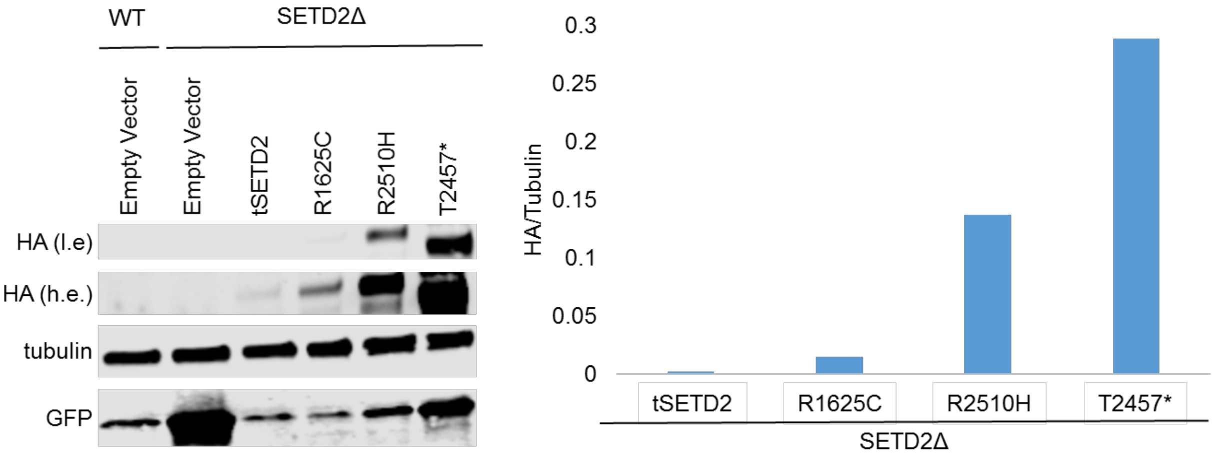


Figure 4.3: SETD2 expression varies between samples. Left: Western for HA tag (l.e. = low exposure, h.e. = high exposure) tubulin loading control and GFP. Right: Quantification of HA levels relative to tubulin.

Both the R2510H SRI point mutation and the T2457* SRI deletion were trimethylation competent, despite predicted disruption of the interaction between tSETD2 and RNA Polymerase II (RNAPII). To further explore the effect of mutation on this relationship, we performed a co-immunoprecipitation for the HA tag and RNAPII. Due to low levels of tSETD2 in 786O stable expressing cells, this was conducted in HKC stable expressing cells, as well as 293T transient expressing cells for verification. As expected, RNAPII co-immunoprecipitated with HA in tSETD2 expressing cells, as well as R1625C expressing cells (Figure 4.5). RNAPII did not co-immunoprecipitate with either R2510H or T2457*, despite both having a similar HA immunoprecipitation efficiency to tSETD2. This was true in both cellular backgrounds.

It is possible that the failure to interact by co-immunoprecipitation is due to the absence of chromatin as a substrate. To test whether chromatin is required for RNAPII to interact with the SRI altered tSETD2,

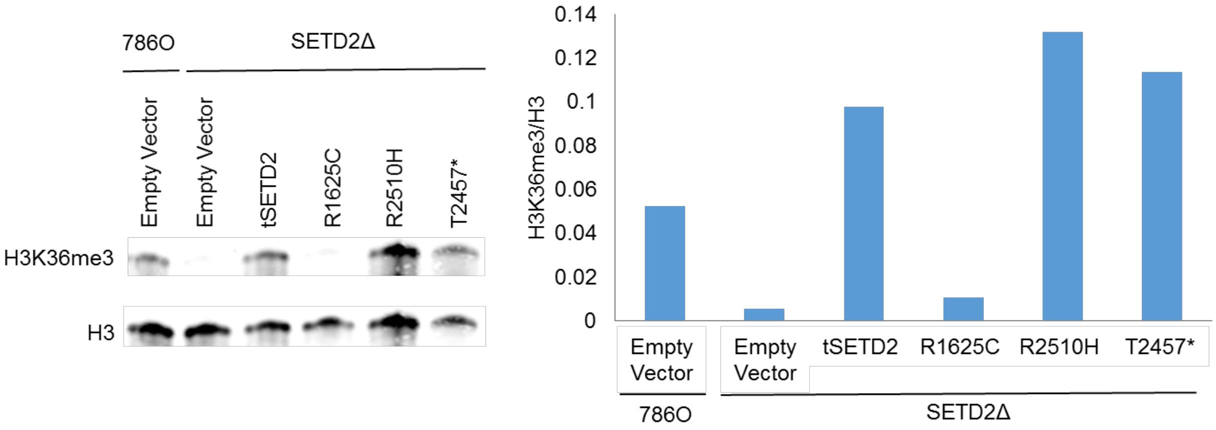


Figure 4.4: SRI domain alteration does not diminish H3K36 trimethylation levels. Left: Histone western blot for H3K36me3 and H3. Right: Quantification of H3K36me3 relative to H3 levels.

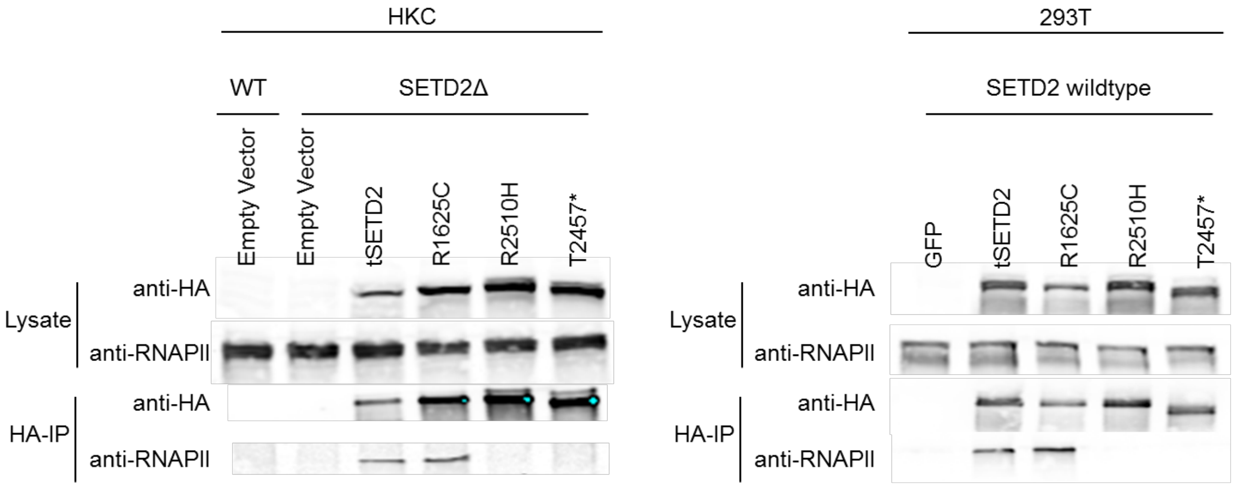


Figure 4.5: SRI domain is required for interaction between tSETD2 and RNAPII. Left: Co-immunoprecipitation in HKC stable expression cells. Right: Co-immunoprecipitation in 293T transient expression cells. Lysate represents whole cell extract used as input for immunoprecipitation. HA-IP represents final sample after immunoprecipitation with anti-HA antibody. Anti-HA and anti-RNAPII refer to the antibody probe used for the western blot.

an HA immunoprecipitation was conducted on a chromatin extract. Unlike the experiments done on cellular extracts, RNAPII co-immunoprecipitated in all samples, including the negative control (Figure 4.6). There was no enrichment for RNAPII in either the tSETD2 or the R1625C samples, despite efficient pulldown of HA. This suggests that the presence of chromatin in the extract does not stabilize the interaction between RNAPII and tSETD2.

4.2.2 ChIP-Rx allows quantitative comparisons between samples

Though the western blot confirms H3K36me3 is present, there is no evidence for or against proper targeting. We hypothesize that as the R2510H and T2457* mutants do not interact with RNAPII (as measured by co-immunoprecipitation), H3K36me3 will be mistargeted in SETD2Δ cells expressing these mutants. This was tested using a modified form of chromatin immunoprecipitation sequencing (ChIP-seq) known as ChIP-Rx [157]. Chromatin Immunoprecipitation with a Reference Exogenous Genome (ChIP-Rx) allows genome-wide quantitative comparisons of histone modification status across cell populations though the use of a reference epigenome. This reference epigenome is one which contains the specific histone modification of interest, and is added equally to each sample. The equal addition of the exogenous genome to each sample permits normalization to the amount of exogenous DNA in the sample, and allows for quantitative comparisons to be made across samples. This technique was chosen over standard ChIP as

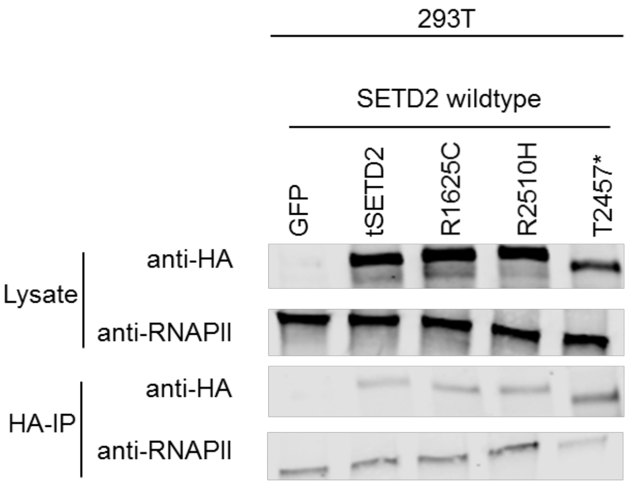


Figure 4.6: Chromatin does not stabilize interaction between tSETD2 and RNAPII. Lysate represents whole cell extract used as input for immunoprecipitation. HA-IP represents final sample after immunoprecipitation with anti-HA antibody. Anti-HA and anti-RNAPII refer to the antibody probe used for the western blot.

differences between SRI altered tSETD2, tSETD2 and wildtype SETD2 may be in terms of levels of H3K36me3.

For ChIP-Rx, the exogenous epigenome chosen was the Drosophila S2 cell epigenome, which also has H3K36me3. S2 nuclei were added at a 1:2 ratio prior to sonication and immunoprecipitation, as this ensures equal sonication and pulldown efficiency for both the sample and the S2 spike-in control (Figure 4.7). After sequencing, the raw files were aligned to both the Drosophila and human genomes to determine the total number of reads for each species. This allows for determination of the normalization constant (α=1/count of reads (in millions) aligning to the Drosophila genome). Comparisons across inputs for each sample shows a similar percentage of reads aligned to the Drosophila genome of the total number of reads (Figure 4.8). This is expected, as the same number of Drosophila cells were added to an identical number of 786O cells. ANOVA testing shows there is no difference between the percent of Drosophila reads across inputs (p=0.935). In H3K36me3 competent samples, the percent of Drosophila reads mapped to each immunoprecipitation sample is similar to the level in the input, though t-test comparison between wildtype input and sample, as well as between tSETD2 input and sample do meet statistical significance (p=0.024

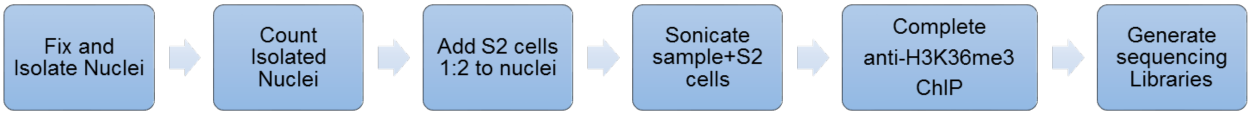

Fix and Isolate Nuclei
Count Isolated Nuclei
Add S2 cells 1:2 to nuclei
Sonicate sample+S2 cells
Complete anti-H3K36me3 ChIP
Generate sequencing Libraries

Figure 4.7: ChIP-Rx uses Drosophila S2 cells as a normalization for H3K36me3 ChIP-seq. Schematic of experimental design for H3K36me3 ChIP-Rx experiment.

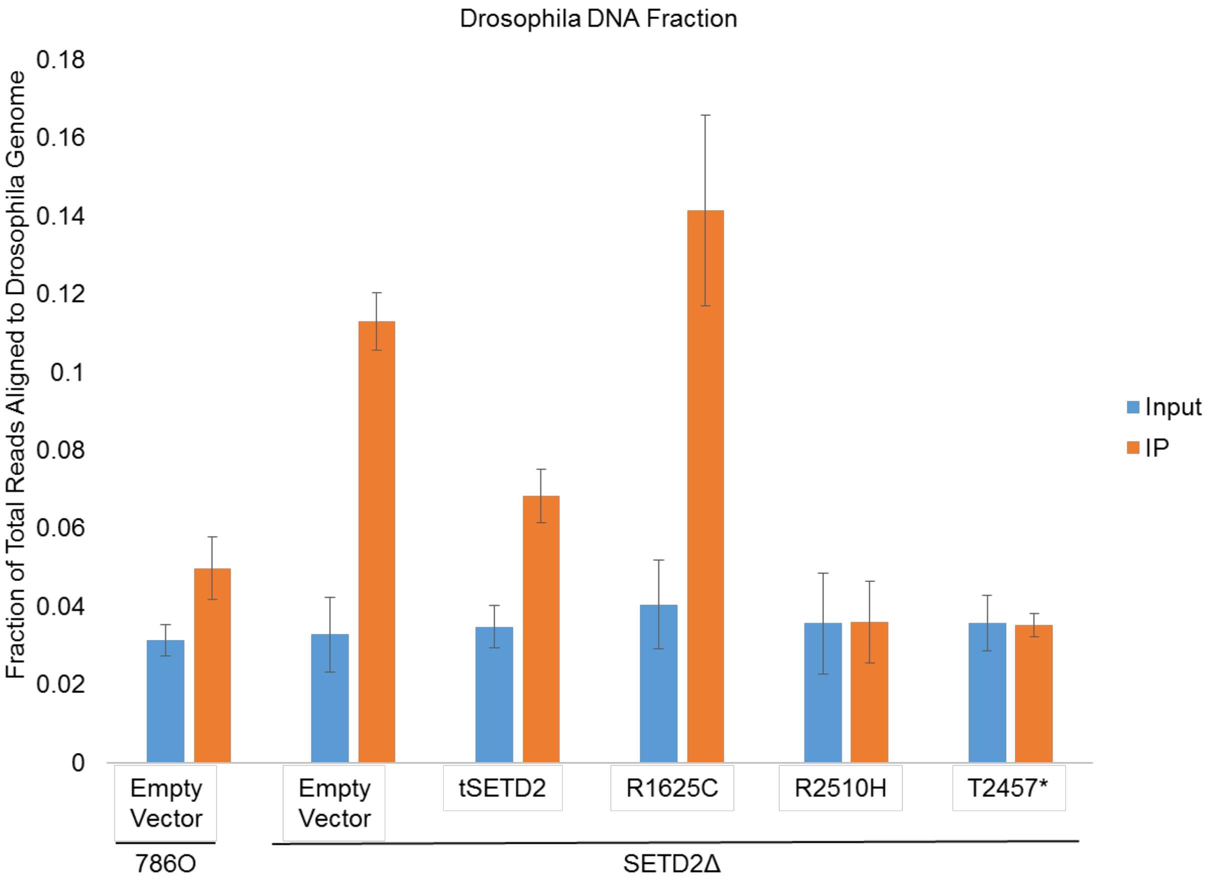


Figure 4.8: H3K36me3 deficient cells show higher levels of Drosophila DNA. Percent of total reads mapping to the Drosophila genome by sample. Input (blue) is the sample prior to immunoprecipitation, IP (orange) is the sample after immunoprecipitation.

and p=0.0090 respectively). Interestingly, H3K36me3 deficient samples show a substantial increase in the percent of reads mapping to Drosophila, with 11.3% of reads in SETD2Δ and 14.1% of reads in SETD2Δ+R1625C mapping to the Drosophila genome (p=0.0022 and p=0.0083 versus input). This is expected, as the Drosophila cells will have the H3K36me3 mark present, and will therefore be enriched over the H3K36me3 deficient SETD2Δ cells in the pulldown.

4.2.3 Wildtype H3K36me3 signal is enriched at coding exons and correlates with gene expression

Previous studies have examined H3K36me3 localization across the genome, however, this is the first large scale H3K36me3 study conducted with a spike-in normalization strategy. This enables high quality analysis of H3K36me3 levels to build upon previously published findings. Initial data analysis focused on genic regions, as H3K36me3 has been linked to transcription [61,158,159]. Metagene analysis

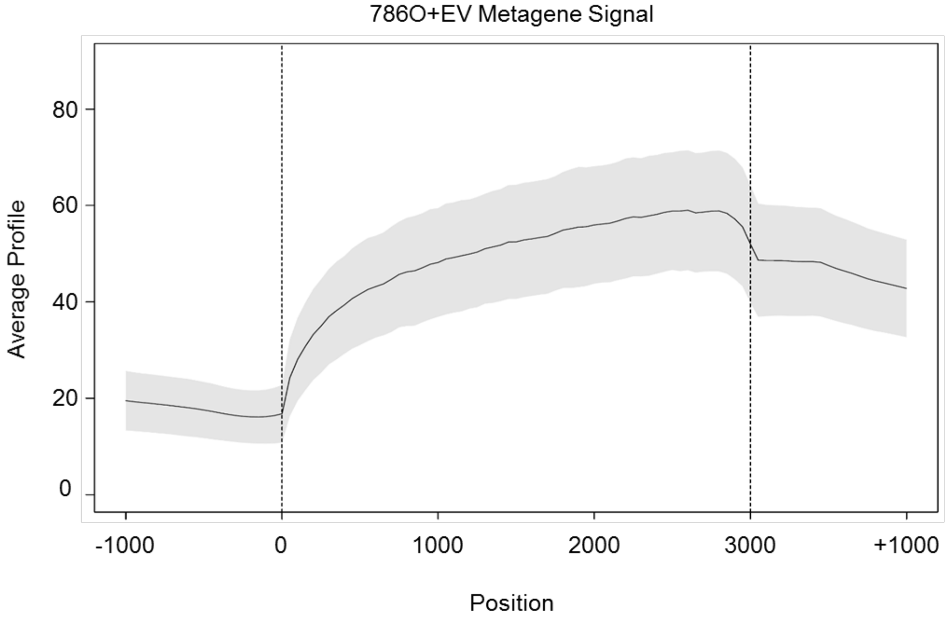


Figure 4.9: 786O + empty vector shows increasing H3K36me3 signal along the length of the gene. Metagene signal of 786O+EV. Average signal (line) for 786O+EV samples 1000 bp upstream of the TSS, from TSS to TTS, and 1000bp downstream of the TTS. Dotted lines represent TSS (left) and TTS (right). Line represents the average signal for all genes, averaged across replicates. Shading represents standard deviation between 786O+EV replicates

of genes shows that H3K36me3 increases along the length of the gene from transcription start site (TSS) to transcription termination site (TTS) (Figure 4.9), similar to previous findings [1,160,161].

When examining signal at specific genomic features, coding exons showed the highest level of signal enrichment compared to input, with the other genic features also showing increased signal relative to input (Figure 4.10). Promoters and intergenic regions showed the least signal enrichment. This result supports previous analyses which showed differential H3K36me3 levels at exons and introns [70], as well as emphasizing the genic localization of H3K36me3.

As signal is most enriched at coding exons, the features of this signal were further explored. Signal at coding exons is not dependent on exon length, as there is not increased signal for longer exons (Figure 4.11). High levels of H3K36me3 are found at long and short exons alike, as are low levels of H3K36me3. The majority of exons are between 30-1000bp in length, and the signal at these exons is shows mostly small differences between sample and input.

In addition to the overall genomic localization of H3K36me3, the effect of gene expression on H3K36me3 levels was examined. RPKM data generated in Chapter 3 was correlated with average H3K36me3 per gene. H3K36me3 levels distributed by RPKM quartiles display a positive correlation, with

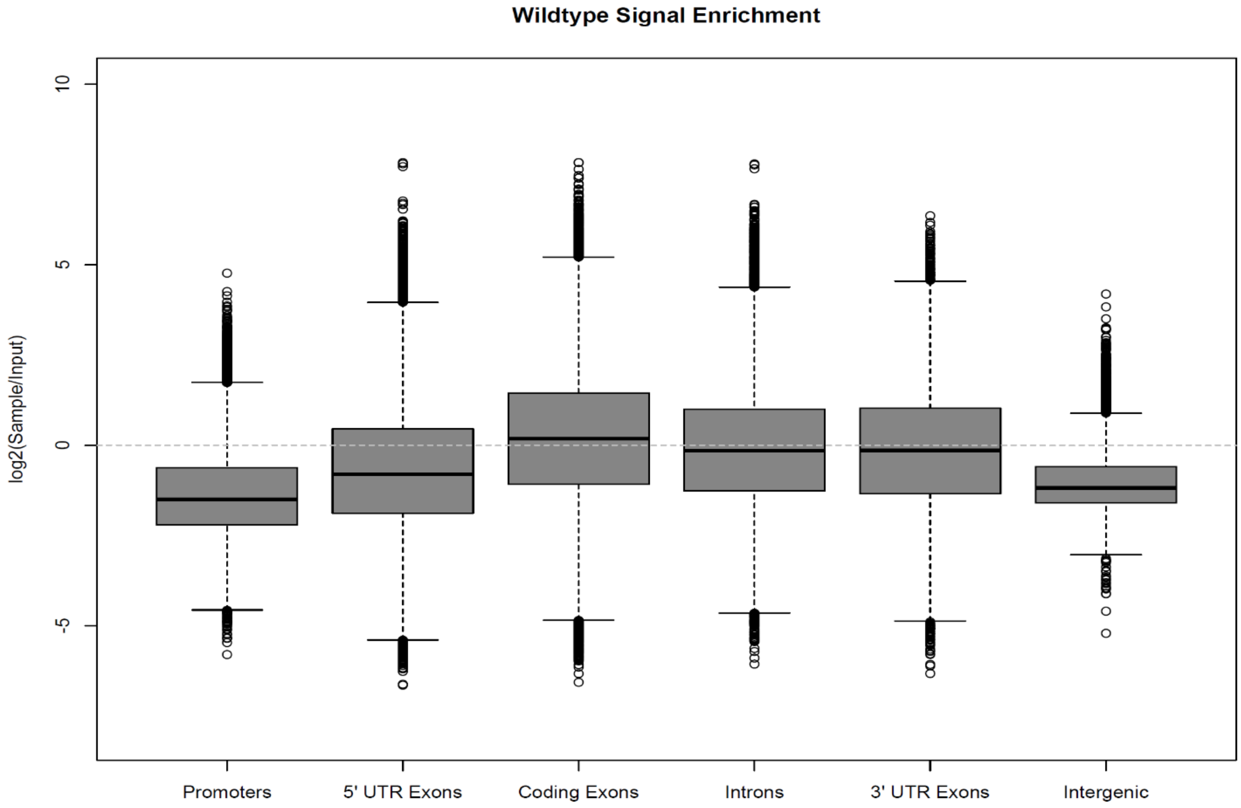


Figure 4.11: Coding Exons show highest level of H3K36me3 enrichment. Boxplots for average signal at each feature. Signal was calculated by dividing H3K36me3 normalized values per feature by the matched input, averaging replicate values, then log2 transformed for plotting. Values greater than 0 (dotted line) are enriched over input.

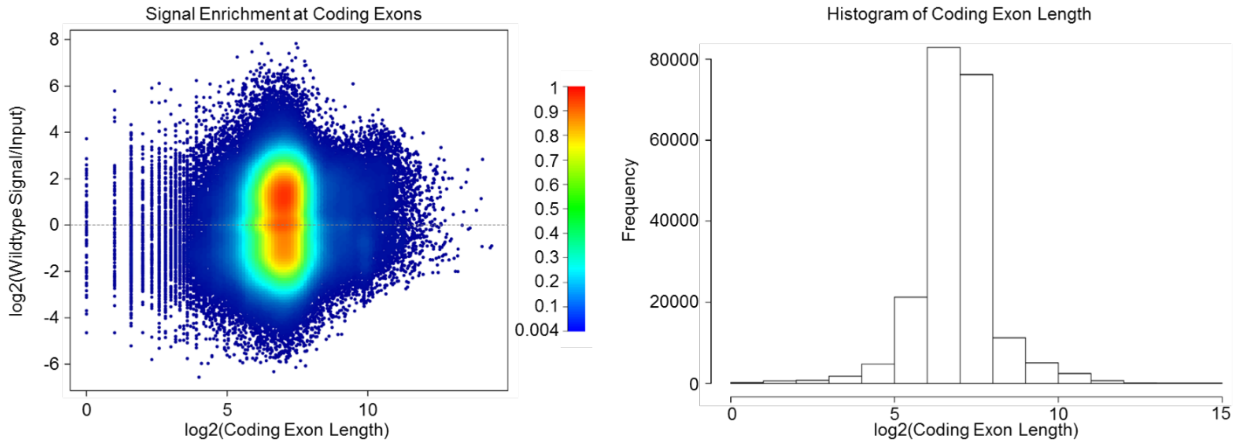


Figure 4.10: H3K36me3 levels are not correlated with exon length. Left: Scatter plot of exon length (log2 transformed) by Sample/Input for 786O+EV samples (log2 transformed). Color indicates density of points at the location, range from blue (low density) to red (high density). Right: Histogram of exon length (log2 transformed).

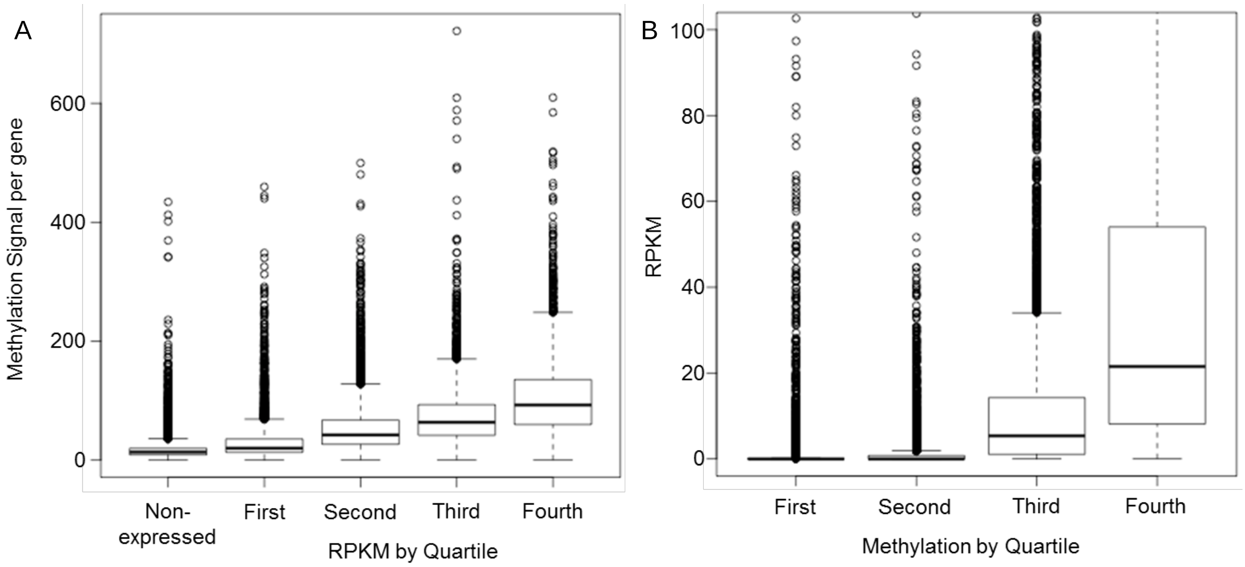


Figure 4.12: RPKM and methylation levels are positively correlated. Left: boxplots showing methylation signal for genes separated by gene RPKM into quartiles. Right: boxplots showing RPKM for genes separated by gene methylation signal into quartiles.

increased gene expression correlating with increased methylation (Figure 4.12A). The reciprocal comparison also shows this trend (Figure 4.12B).

4.2.4 SETD2Δ results in total loss of H3K36me3

The benefit of ChIP-Rx is the normalization factor provided by the Drosophila cells. The addition of cells which have the positive mark decreases the likelihood of finding false positives in the ChIP-seq dataset for SETD2Δ cells. Based on this assumption, H3K36me3 ChIP was conducted in the SETD2Δ cells to determine if loss of SETD2 results in retargeting of trimethylation to non-coding regions. We first examined signal at genic regions, as was done for 786O+EV samples. Unlike the SETD2 wildtype cells, SETD2Δ+EV cells show near complete loss of signal at genes (Figure 4.13A).

When examining signal at specific genomic features, the generalized loss of H3K36me3 signal can be clearly observed (Figure 4.13B). Nearly all the data has a log2 ratio of Sample/Input less than 0, indicating that few data points show higher signal in the H3K36me3 ChIP than the matched input. Comparing signal specifically at coding exons shows very little signal greater than zero in the SETD2Δ+EV samples, regardless of whether signal is enriched above input levels in SETD2 wildtype cells (Figure 4.13C). Overall, this supports the role of SETD2 as a non-redundant H3K36 trimethyltransferase in human cells.

4.2.5 Expression of R1625C SET domain mutant fails to restore trimethylation to H3K36 across the genome

As SETD2Δ cells lack H3K36me genome-wide and the R1625C mutant was previously shown to be catalytically inactive (Chapter 2, [152]), we examined H3K36me3 levels across the genome in SETD2Δ cells stably expressing the R1625C mutant. We hypothesized that there would be little change observed between these samples and SETD2Δ samples. Indeed, examining genes shows little to no signal present in the SETD2Δ+R1625C sample (Figure 4.14A). Across genomic features, there are few features which

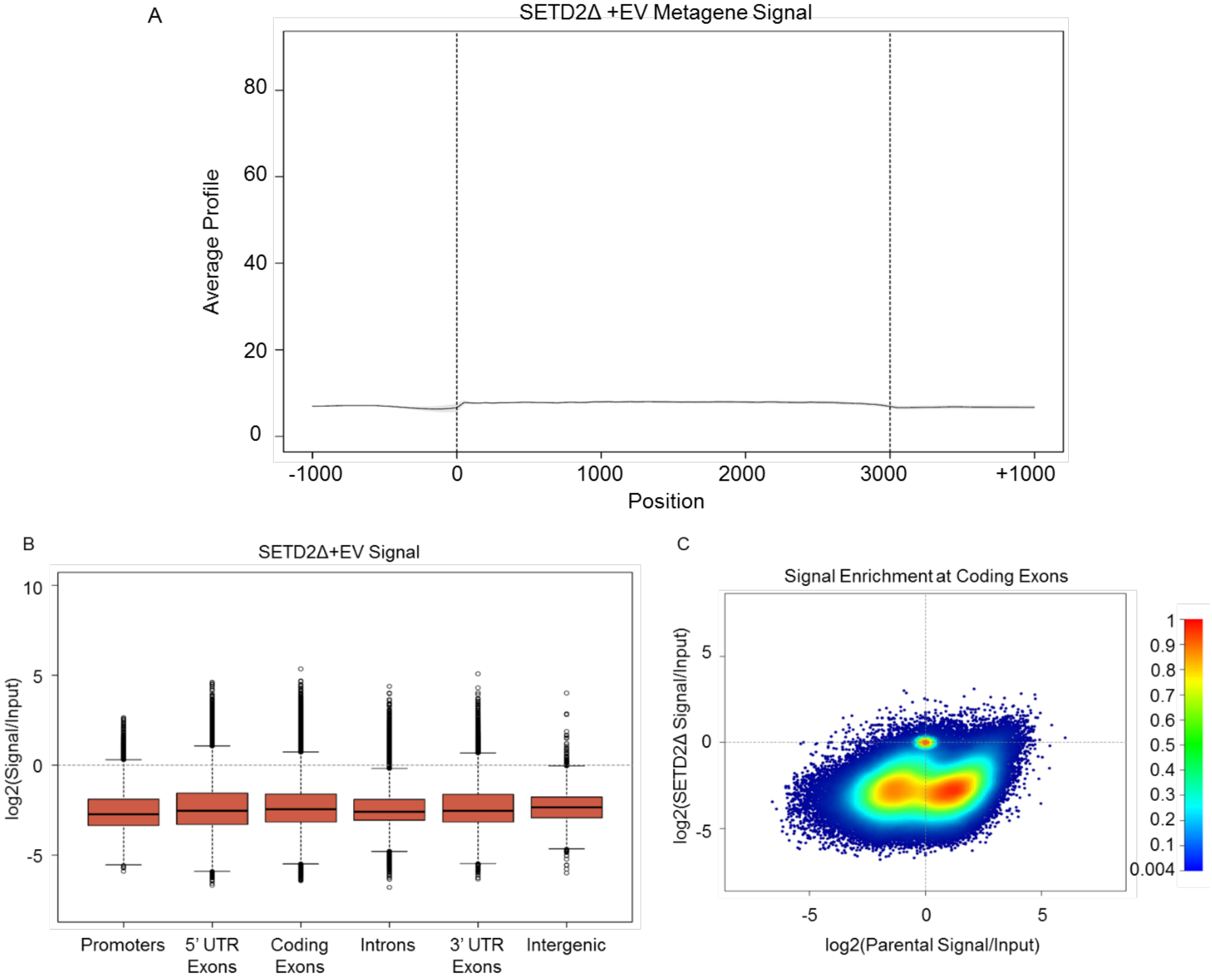


Figure 4.13: SETD2 loss results in loss of H3K36me3 across the genome. A) Metagene signal of SETD2Δ+EV. Average signal (line) for SETD2Δ+EV samples 1000 bp upstream of the TSS, from TSS to TTS, and 1000bp downstream of the TTS. Dotted lines represent TSS (left) and TTS (right). B) Boxplots for average signal at each feature. Signal was calculated by dividing H3K36me3 normalized values per feature by the matched input, averaging replicate values, then log2 transformed for plotting. Values greater than 0 (dotted line) are enriched over input. C) Signal at coding exons comparing data from Figure 4.10 and 4.13B. Color indicates density of points at the location, range from blue (low density) to red (high density).

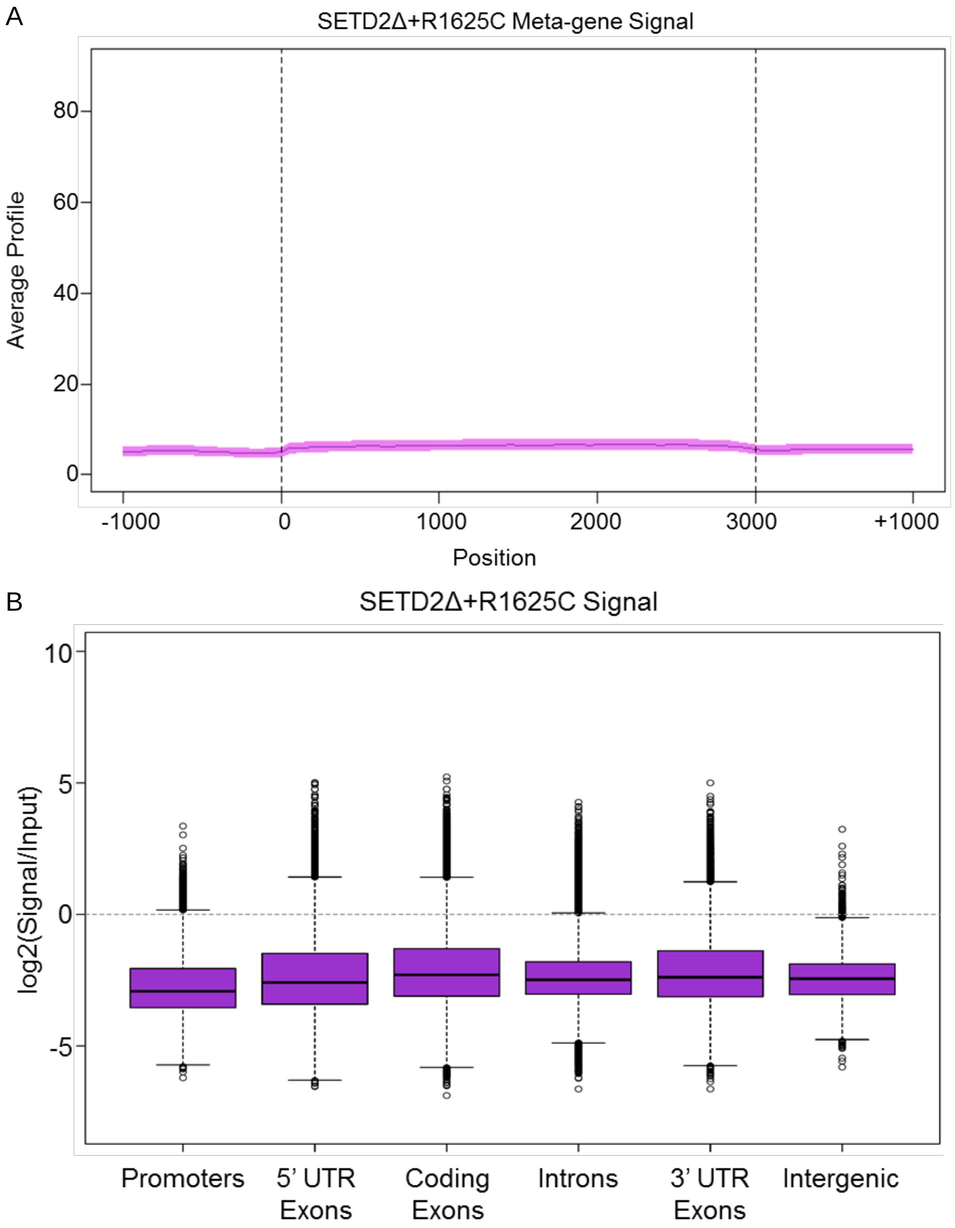


Figure 4.14: SET domain mutation shows little to no signal across genomic features. A) Metagene signal of SETD2Δ+R1625C samples. Average signal (line) for SETD2Δ samples 1000 bp upstream of the TSS, from TSS to TTS, and 1000bp downstream of the TTS. Dotted lines represent TSS (left) and TTS (right). Line represents the average signal for all genes, averaged across replicates. Shading represents standard deviation between replicates B) Boxplots for average signal at each feature. Signal was calculated by dividing H3K36me3 normalized values per feature by the matched input, averaging replicate values, then log2 transformed for plotting. Values greater than 0 (dotted line) are enriched over input.

display signal levels greater than input, similar to what was seen in SETD2Δ samples (Figure 4.14B). This data supports the characterization of the R1625C mutant as catalytically inactive.

4.2.6 H3K36me3 positive and H3K36me3 negative samples are distinct populations

As SETD2Δ+EV and SETD2Δ+R1625C cells have been shown to be lack H3K36me3, we explored the differences between these H3K36me3 negative cells and the H3K36me3 positive cells. First, principal component analysis was conducted on the immunoprecipitation samples and the matched inputs. The first component correlates with H3K36me3 status, as the positive samples (SETD2 wildtype, SETD2Δ+tSETD2, SETD2Δ+R2510H, SETD2Δ+T2457*) span PC1 (Figure 4.15A). The H3K36me3 negative samples cluster with the input samples in PC1.

The inputs were removed from analysis, and the principal component analysis was repeated. PC1 again correlates with H3K36me3 status (Figure 4.15B). Each sample type clusters most closely with its replicates, suggesting some variation between backgrounds. The exception to this trend is SETD2Δ+EV and SETD2Δ+R1625C, which cluster with each other, suggesting high levels of similarity between these samples. PC2 in this analysis separates the 786O+EV samples from all others, which may indicate some variation based on the SETD2Δ background.

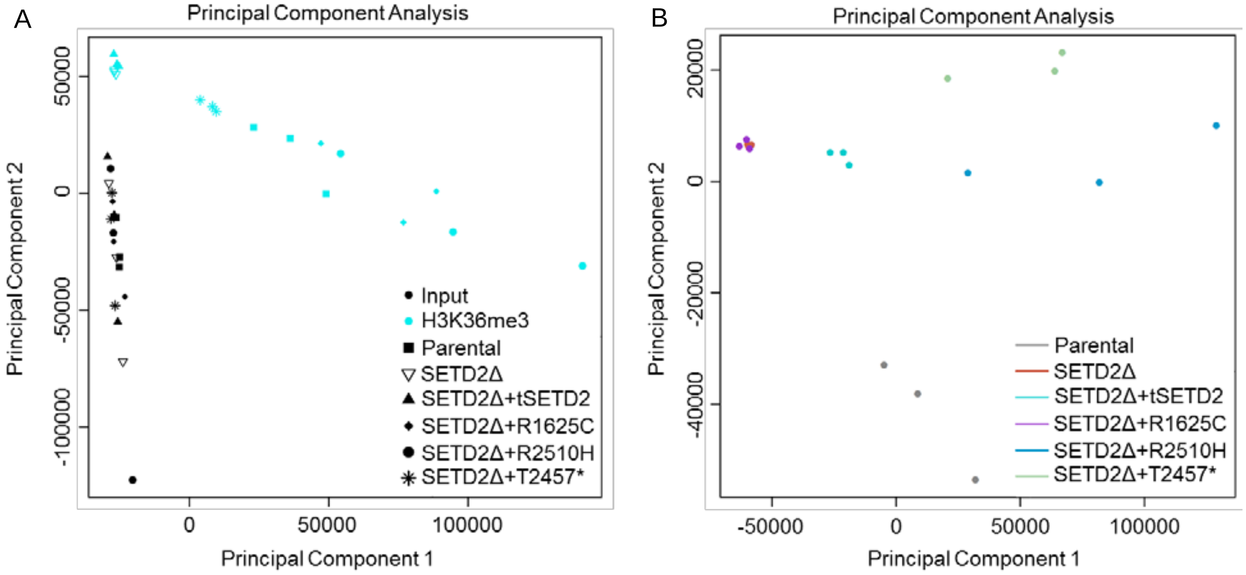


Figure 4.15: Principal Component Analysis separates samples based on H3K36me3 status. A) PCA of input and IP samples. Input samples shown in black, IP in cyan. Shape denotes sample type. B) PCA of IP samples only. Color denotes sample type.

For further examination of the differences between H3K36me3 positive and negative samples, we determined the average signal in 500 base pair windows across the genome, and used DESeq2 [141] to cluster this data. We found that H3K36me3 positive and H3K36me3 negative samples form two distinct clusters (Figure 4.16A). This data also shows 786O+EV and SETD2Δ+tSETD2 samples clustering more closely together than to SRI altered samples, though the sample distances are similar. An additional clustering algorithm, ConsensusClusterPlus [162], generated a similar result (Figure 4.16B). In addition to clustering results, conducting multiple unpaired t-tests between samples using 500 base pair windows and controlling for false discovery rate (α<0.05) using the Benjamini-Hochberg method [163] allows determination of the number of windows which reject the null hypothesis (that there is no difference between samples). The comparisons were made pairwise between samples. The number of significant windows can be used to generate a dendogram. This analysis shows that SETD2Δ+EV and SETD2Δ+R1625C separate from the H3K36me3 positive, and have large numbers of 500bp windows which are significantly different from the other samples (Figure 4.16C). Based on these analyses, SETD2Δ+EV and SETD2Δ+R1625C appear to be fundamentally distinct from the other H3K36me3 positive samples.

4.2.7 SRI altered tSETD2 shows few differences to wildtype SETD2

As SETD2Δ cells lack H3K36me3, comparisons between tSETD2, R2510H and T2457* samples with wildtype will be based entirely on the function of the stably expressed construct. As with the 786O+EV samples, the signal at genes was examined using a metagene plot. SETD2Δ+tSETD2, SETD2Δ+R2510H, and SETD2Δ+T2457* all restore H3K36me3 in a similar distribution to wildtype (Figure 4.17). Additionally, examination of signal at specific genomic features reveals a similar pattern of enrichment to that seen in 786O+EV samples (Figure 4.18). Coding exons show the higher level of enrichment over input in all H3K36me3 positive samples, while intergenic regions and promoters show very little enrichment. This suggests that SRI domain alteration through mutation or deletion do not drastically alter the overall positioning of H3K36me3 across the genome.

To compare signal at specific genomic regions, the Benjamini-Hochberg method was used again. As before, a dendogram was created for each region. At all regions examined, SETD2Δ+EV and SETD2Δ+R1625C fell on one arm of the dendogram (Figure 4.19), supporting results described above. Also as above, H3K36me3 positive samples consistently group together. 786O+EV, SETD2Δ+R2510H and

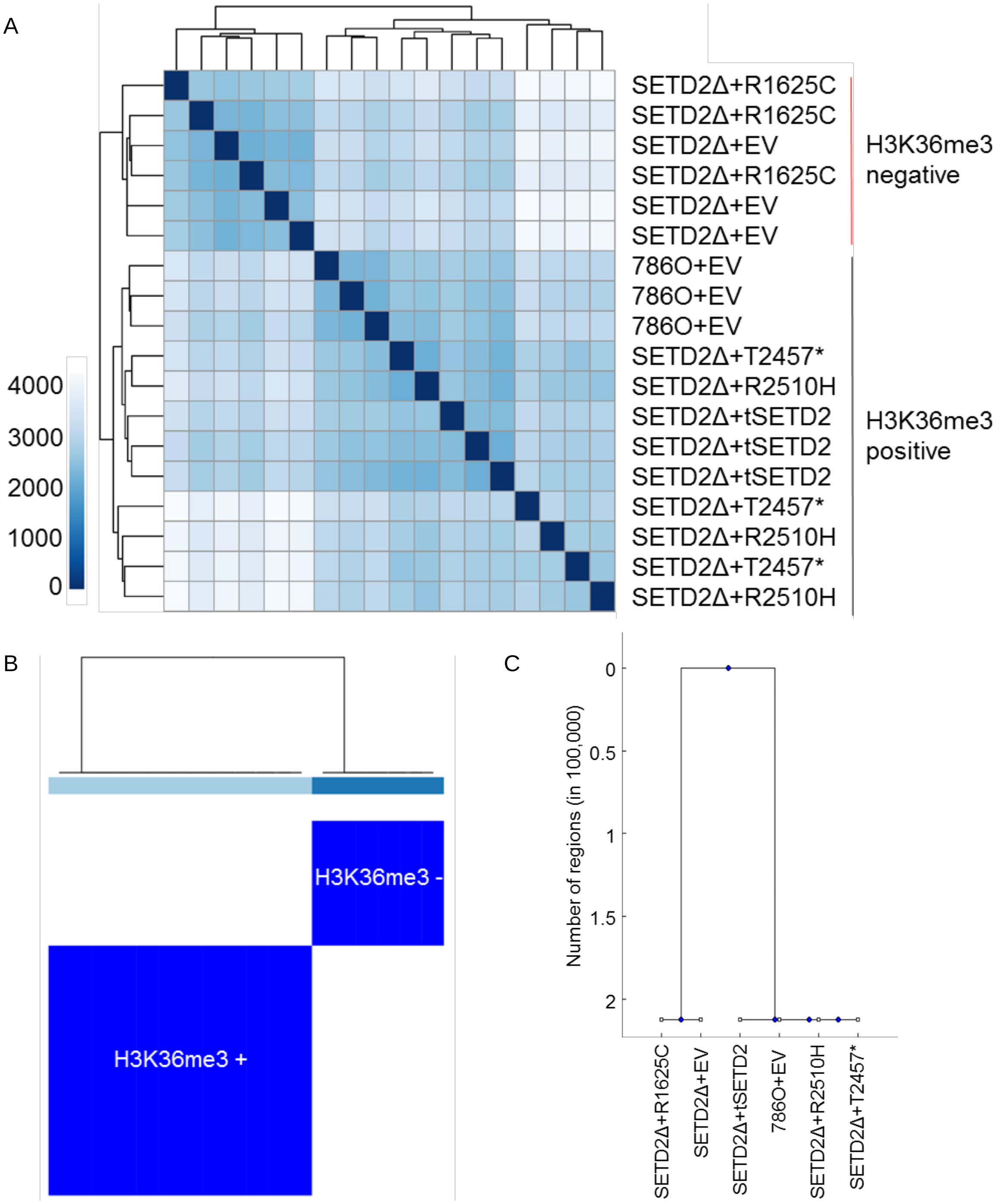
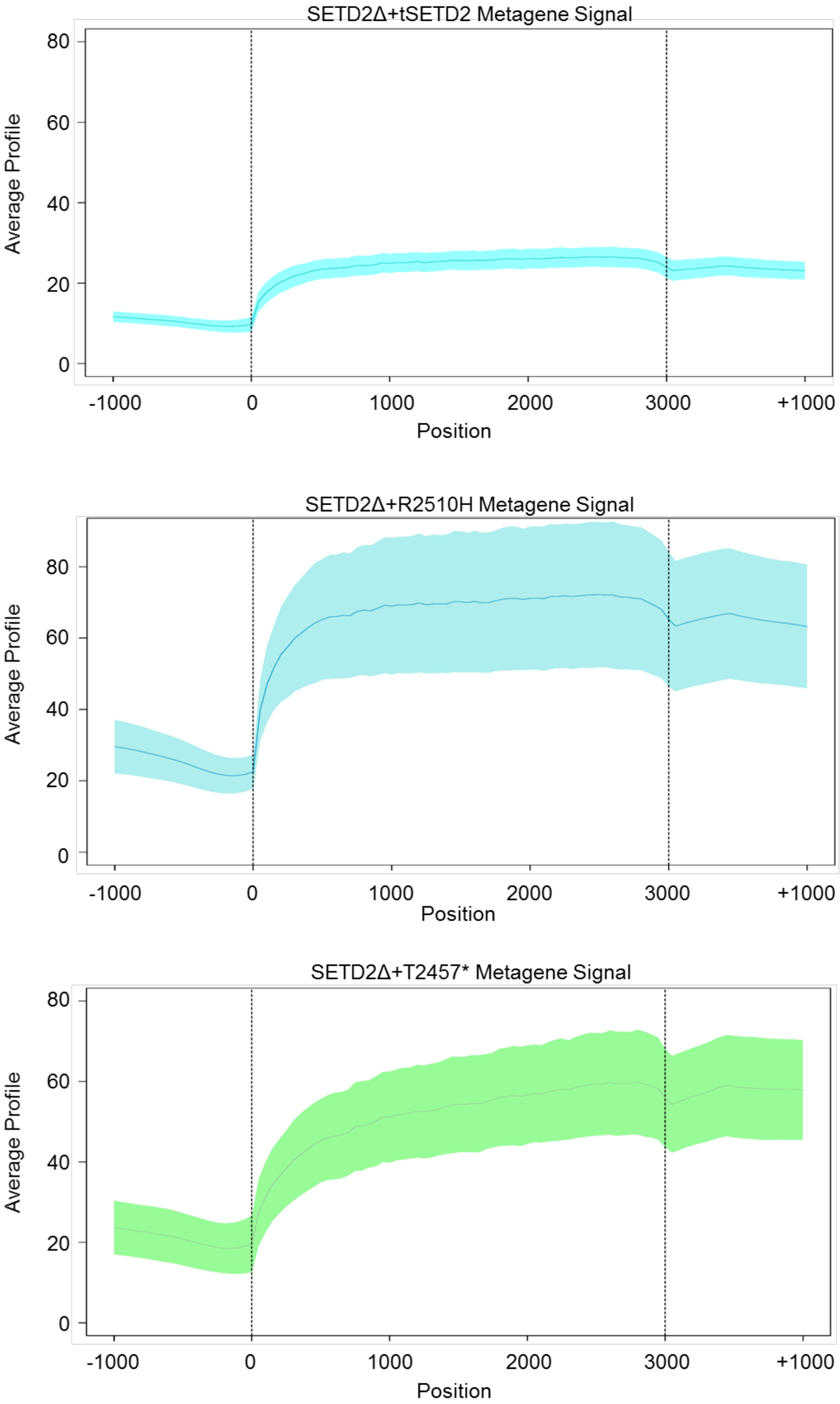


Figure 4.16: H3K36me3 positive and H3K36me3 negative samples separate based on clustering of 500bp windows. A) Cluster generated using DESeq2 analysis of 500bp windows B) Cluster generated using ConsensusClusterPlus analysis of 500bp windows. C) Dendogram created based on number of 500bp windows which reject the null hypothesis when compared to 786O+EV in the Benjamini-Hochberg method. Y-axis represents the number of regions between samples.



SETD2Δ+tSETD2 Metagene Signal

SETD2Δ+R2510H Metagene Signal

SETD2Δ+T2457* Metagene Signal

Figure 4.17: tSETD2, R2510H, and T2457* restore H3K36me3 to genes in a similar pattern to wildtype SETD2. Top: SETD2Δ+tSETD2. Middle: SETD2Δ+R2510H. Bottom: SETD2Δ+T2457*. Metagene signal of SETD2Δ+R1625C samples. Average signal (line) for 786O+EV samples 1000 bp upstream of the TSS, from TSS to TTS, and 1000bp downstream of the TTS. Dotted lines represent TSS (left) and TTS (right). Line represents the average signal for all genes, averaged across replicates. Shading represents standard deviation between replicate.

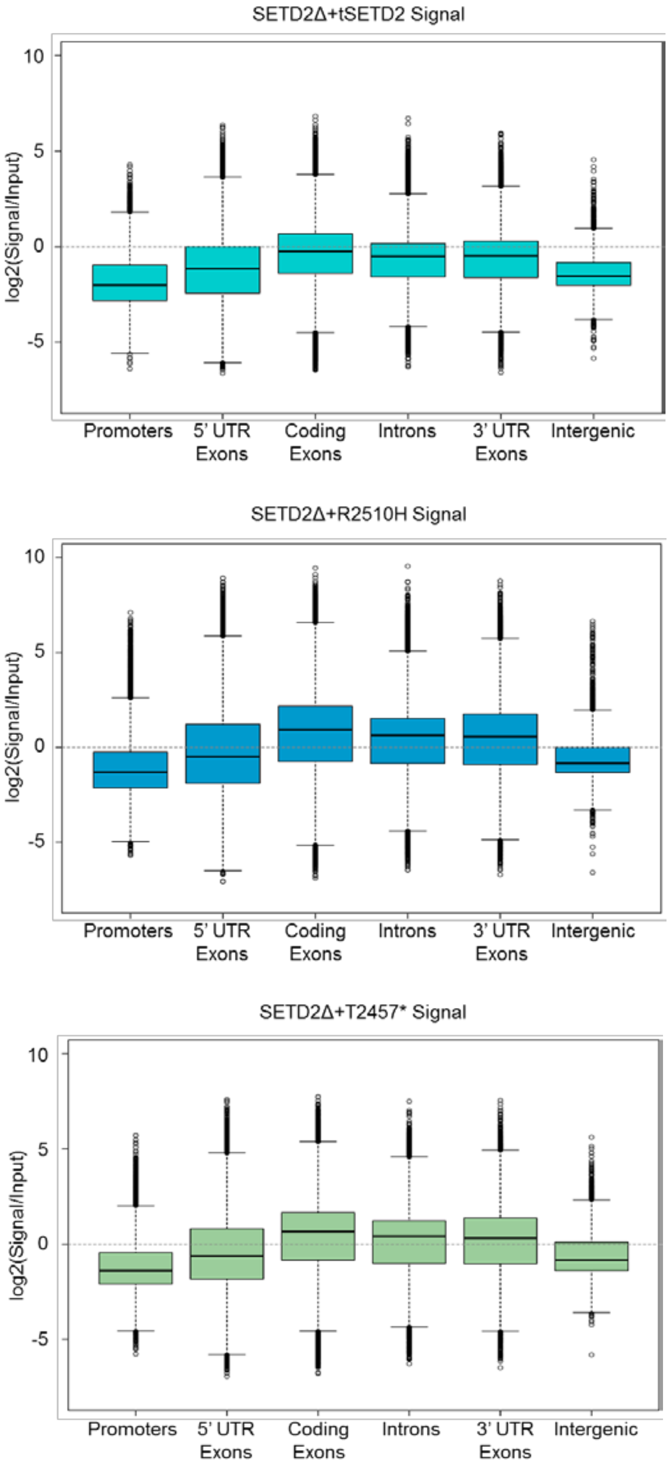


Figure 4.18: Coding exons show highest enrichment level across saomples. Boxplots for average signal at each feature. Signal was calculated by dividing H3K36me3 normalized values per feature by the matched input, averaging replicate values, then log2 transformed for plotting. Values greater than 0 (dotted line) are enriched over input. Top: SETD2Δ+tSETD2. Middle: SETD2Δ+R2510H. Bottom: SETD2Δ+T2457*

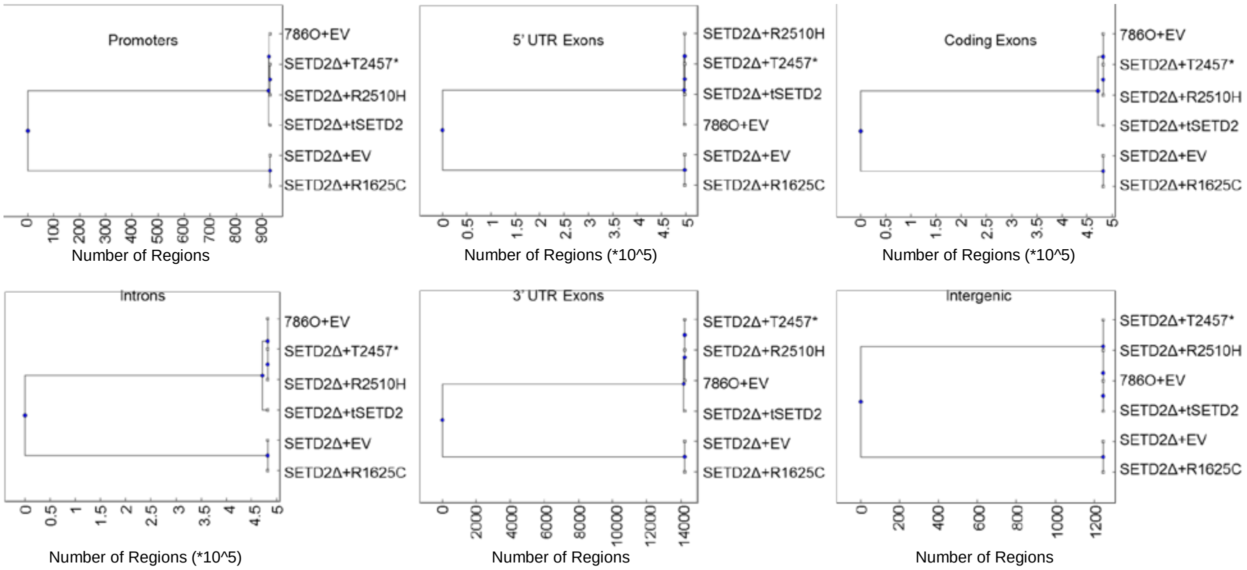


Figure 4.19: Benjamini-Hochberg method identifies few differential regions between H3K36me3 positive samples. Dendogram created based on number of the given region which reject the null hypothesis when compared to 786O+EV in the Benjamini-Hochberg method. Dendograms were created for each examined genomic feature. X-axis represents the number of significant regions between samples.

SETD2Δ+T2457* show very little differences at any of the genomic features. As coding exons are the most enriched feature of wildtype samples, these were examined more in depth. The number of coding exons which reject the null hypothesis varies between samples. Compared to 786O+EV, less than 20 exons reject the null hypothesis for both SETD2Δ+R2510H and SETD2Δ+T2457* (Table 4.1). This is in stark contrast to the ~90,000 exons which reject the null hypothesis for 786O+EV compared to SETD2Δ+EV and SETD2Δ+R1625C. Interestingly, comparing 786O+EV and SETD2Δ+tSETD2 shows an intermediate phenotype, with ~24,000 exons that reject the null hypothesis. Overall, these comparisons suggest that there are very few differences between SRI altered tSETD2 and wildtype SETD2 in terms of H3K36me3 deposition.

	786O +EV	**SETD2Δ +EV**	**SETD2Δ +tSETD2**	**SETD2Δ +R1625C**	**SETD2Δ +R2510H**	**SETD2Δ +T2457***
786O +EV	0	89047	23607	95099	4	16
SETD2Δ +EV		0	48178	10	71350	77301
SETD2Δ +tSETD2			0	56167	1004	2388
SETD2Δ +R1625C				0	76292	82430
SETD2Δ +R1625C					0	0
SETD2Δ +T2457*						0

Table 4.1: SRI altered tSETD2 resembles wildtype SETD2 at coding exons. Number of Coding Exons which reject the null hypothesis of no difference between the row name and column name using the Benjamini-Hochberg method.

4.2.8 T2457* samples have an altered H3K36me3 pattern at the TSS

As the metagene shows a rapid increase from the TSS into the gene, we examined H3K36me3 signal near the TSS. In 786O+EV samples, H3K36me3 increases from the TSS into the gene, similar to the result seen in the metagene plot (Figure 4.20A). We next examined the role gene expression plays on this phenotype. Genes were separated by RPKM into quartiles and H3K36me3 methylation signal at the TSS was plotted for the first and fourth quartile of gene expression (Figure 4.20B). Genes in the first quartile of gene expression showed a similar, if diminished, pattern to that seen in all genes, with signal increasing from the TSS into the gene. Genes in the fourth quartile showed a large increase in overall signal level, indicating again the link between H3K36me3 levels and gene expression.

As there is a clear pattern for H3K36me3 signal at the TSS, the signal in the other H3K36me3 positive samples was examined. SETD2Δ+tSETD2 samples showed diminished signal, but the same

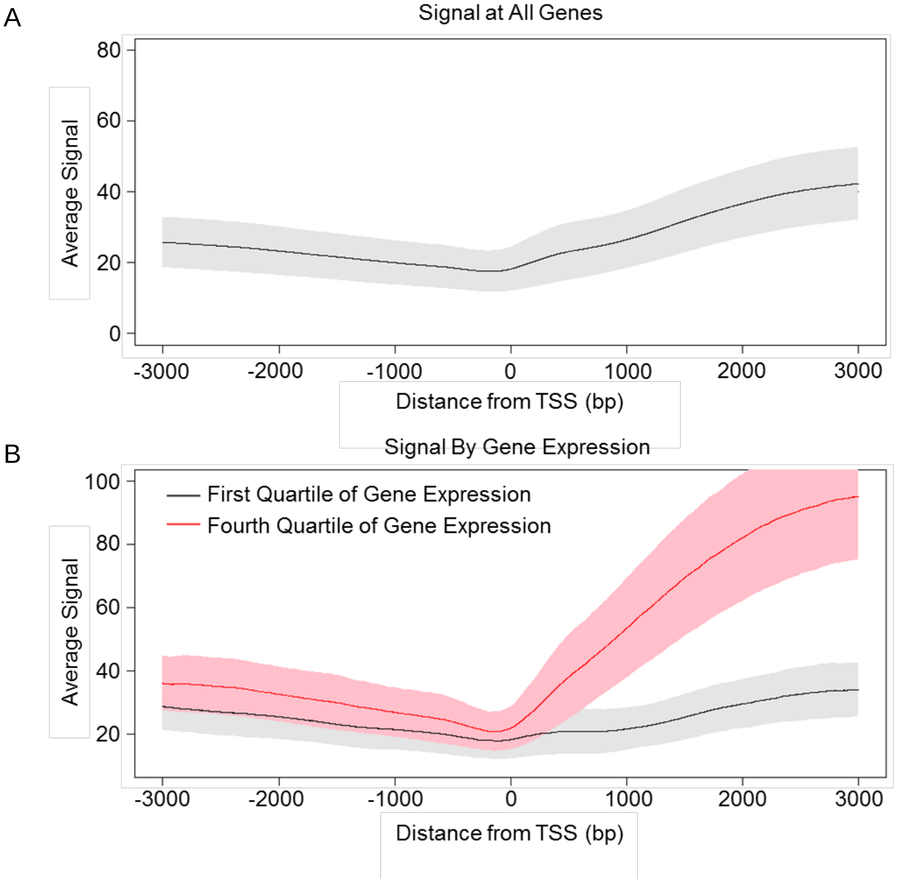


Figure 4.20: H3K36me3 signal increases from the TSS into the gene and correlates with gene expression. A) H3K36me3 signal at all genes B) H3K36me3 signal for genes in the first and fourth quartile of expression by RPKM. Average signal (line) for 786O+EV samples 1500 bp upstream of the TSS and 1500bp downstream of the TSS. Line represents the average signal for all genes, averaged across replicates. Shading represents standard deviation between replicate

overall pattern as 786O+EV (Figure 4.21A). SETD2Δ+R2510H samples had higher levels of variation, but again showed the same general pattern as 786O+EV. An interesting phenomenon was observed in the SETD2Δ+T2457*. Rather than a smooth transition from the TSS into the gene, there is an increase in signal approximately 500bp downstream of the TSS (Figure 4.21B). This point of inflection is present regardless of gene expression levels (Figure 4.21C).

To further explore this phenotype, we examined individual genes which displayed the point of infection in signal. These genes were defined in two ways. First, a subtraction model was used. This model subtracts the signal of SETD2Δ+tSETD2 from SETD2Δ+T2457* signal to find genes in which: a) had a Poisson peak present in the SETD2Δ+T2457* signal, b) subtraction signal which exceeded a threshold of α = 10E-5 (positive or negative) and c) have a region of positive threshold-exceeding subtraction signal upstream of the region of negative threshold-exceeding subtraction signal. This method identified 519 genes. The second method used was a principal component analysis. When examining the signal for SETD2Δ+T2457* samples at the TSS, the signal can be broken into principal components using singular-value decomposition. The first component (PC1) for SETD2Δ+T2457* and SETD2Δ+tSETD2 is the same, and represents the signal increase from the TSS into the gene (Figure 4.22). The second component (PC2) identifies the peak that can be seen as the major signal. Those genes which had PC2 values above the 95th percentile were selected, which created a final list of 170 genes. The genes used for both analyses were those in the top quartile of gene expression by RPKM. The intersection between these two very stringent methods resulted in a final list of 11 genes.

These genes were examined in the UCSC Genome Browser [164], and the upstream bias of signal is stark (Figure 4.23). All genes showed increased signal immediately at the TSS in SETD2Δ+T2457* compared to 786O+EV and SETD2Δ+tSETD2. Surprisingly, the upstream bias is also found in the SETD2Δ+R2510H sample. The SETD2Δ+T2457* and SETD2Δ+R2510H most closely resemble each other at these genes, suggesting this phenotype is due to the lost interaction with RNAPII.

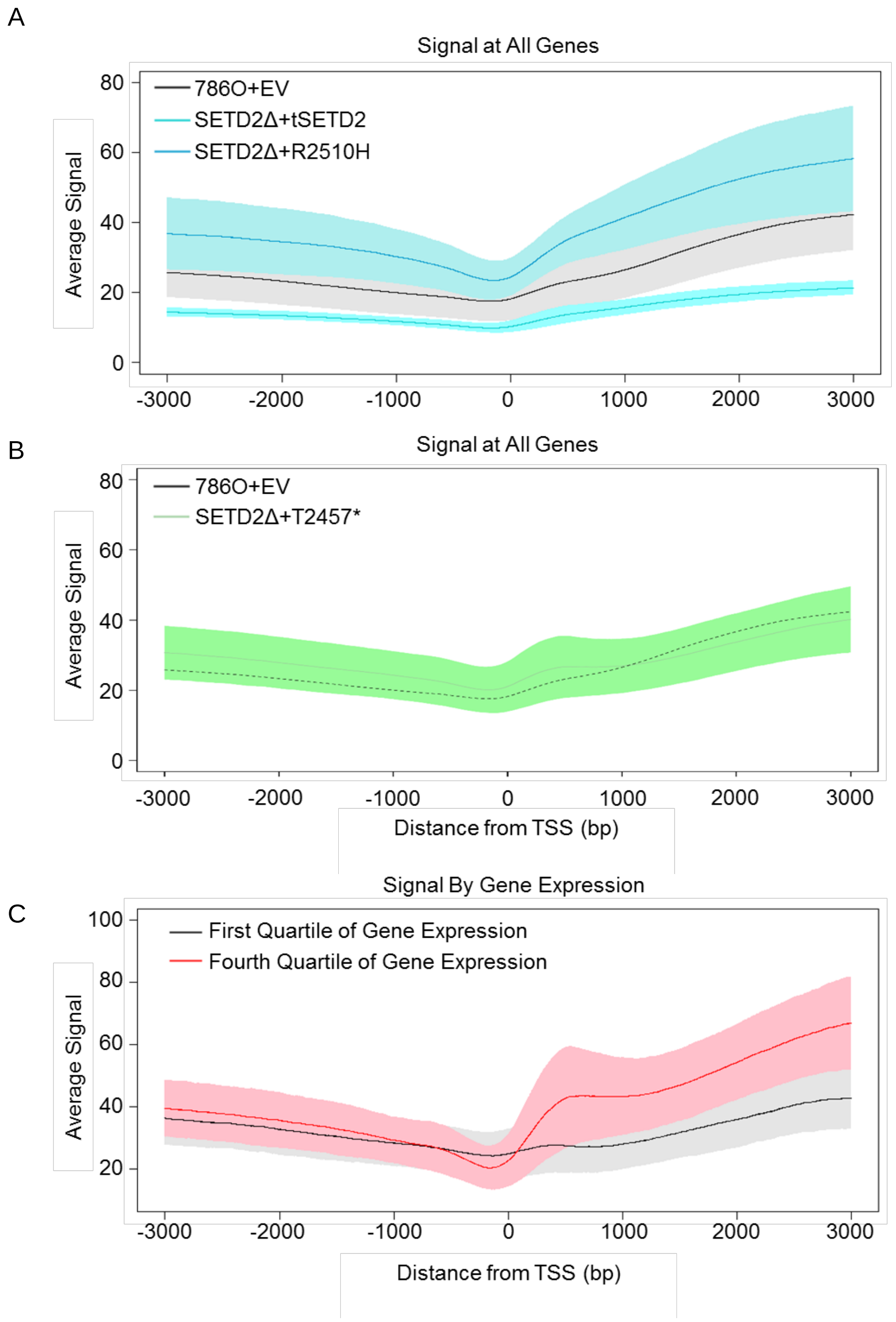


Figure 4.21: SETD2Δ+T2457* signal at the TSS is altered from 786O+EV signal. Average signal (line) for samples 1500 bp upstream of the TSS and 1500bp downstream of the TSS. Line represents the average signal for all genes, averaged across replicates. Shading represents standard deviation between replicate A) H3K36me3 signal at all genes for 786O+EV, SETD2Δ+tSETD2 and SETD2Δ+R2510H samples. B) H3K36me3 signal for SETD2Δ+T2457* samples with dotted line representing 786O+EV average signal C) H3K36me3 signal for genes in the first and fourth quartile of expression by RPKM.

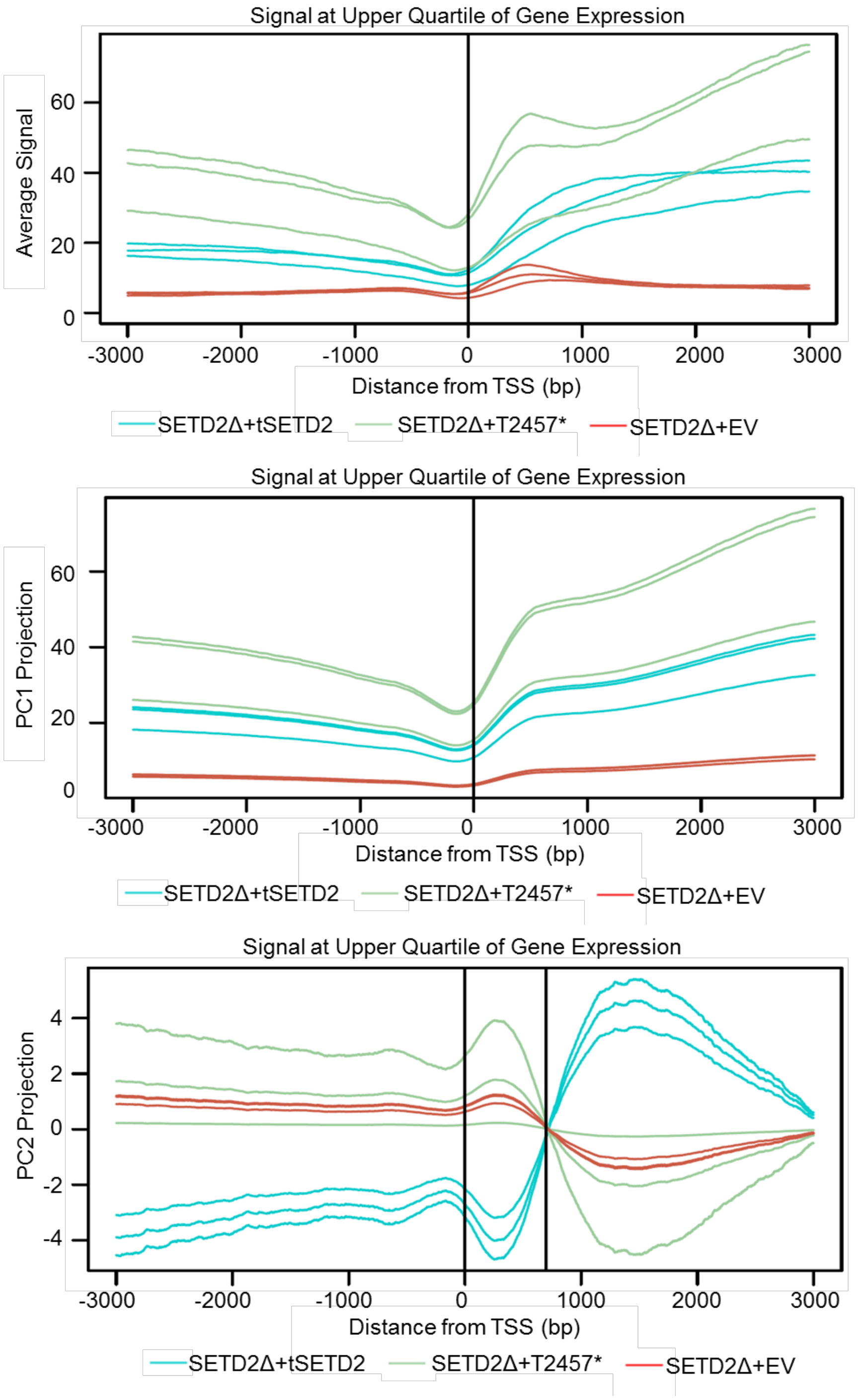


Figure 4.22: Principal component analysis show presence of inflection point in SETD2Δ+T2457*. Signal at the TSS for SETD2Δ+tSETD2, SETD2Δ+T2457* and SETD2Δ+EV. Signal is plotted per replicate. Top: H3K36me3 signal at genes in the top quartile of gene expression. Middle: principal component one of the signal from top. Bottom: principal component two of signal from top. Black vertical lines represent region used to identify genes with point of inflection in SETD2Δ+T2457* samples.

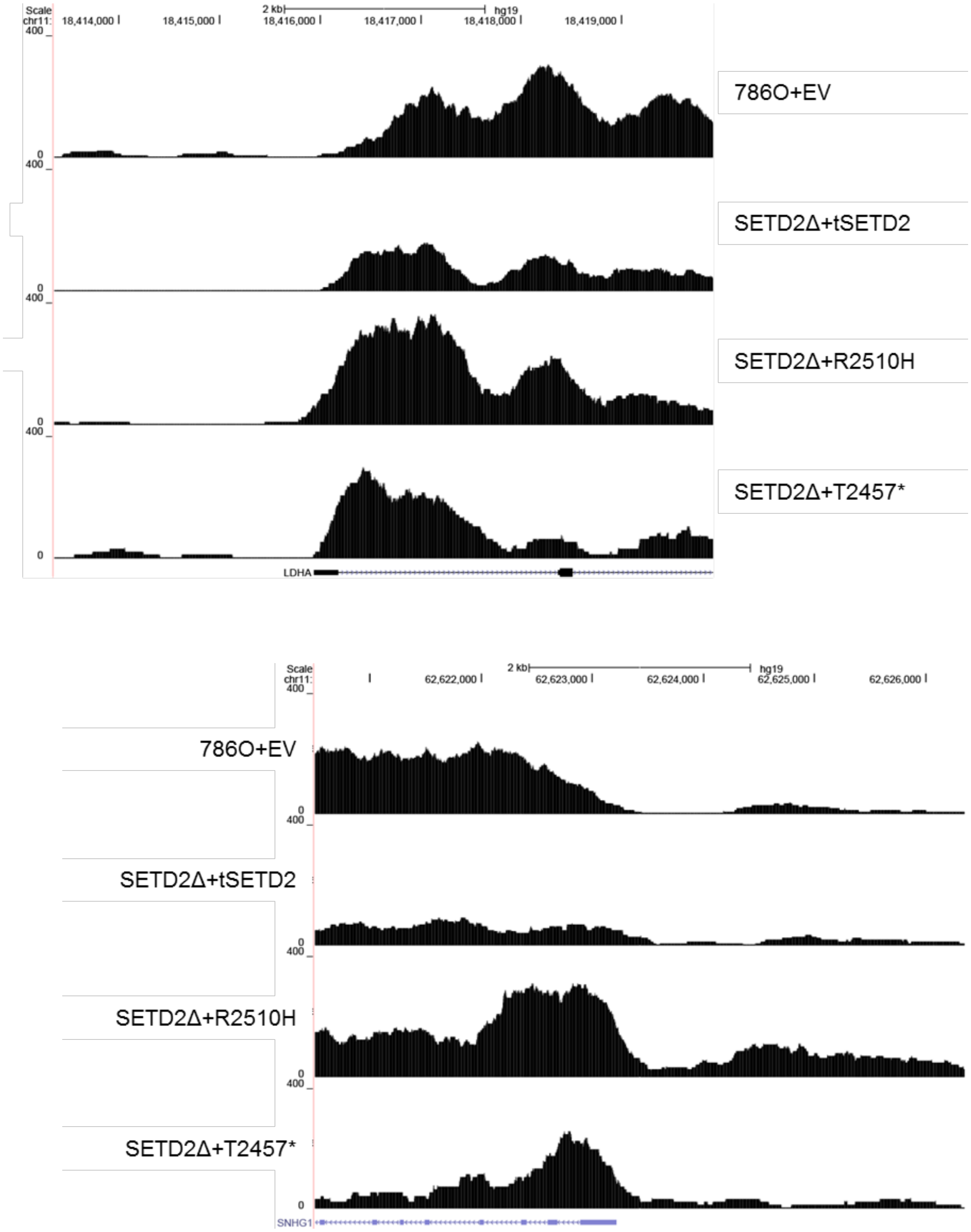


Figure 4.23: H3K36me3 signal at individual genes shows upstream bias of SRI altered samples. Sequencing tracks obtained from the UCSC genome browser [164]. Top: H3K36me3 signal at the LDHA locus. Y-axis is signal levels, x-axis is genomic position. LDHA starts in center and proceeds to the right. Bottom: H3K36me3 signal at the SNHG1 locus. Y-axis is signal levels, x-axis is genomic position. SNHG1 starts in center and proceeds to the left. LDHA and SNHG1 are two representative genes of this phenomenon.

4.3 Discussion

In an effort to understand the effect of mutation on H3K36me3 positioning, we conducted ChIP-Rx for this mark in a variety of SETD2 mutant backgrounds. In this chapter, we demonstrated complete loss of H3K36me3 in SETD2Δ cells, as well as verified the catalytic inactivation of SETD2 by the R1625C mutation. This is in contrast to previously published work which showed that loss of SETD2 through zinc-finger nuclease inactivation results in gains in H3K36me3 in intergenic regions [94]. This study was also conducted in 786O, so genomic background is not the source for the difference. The study, however, did not use a spike-in normalization strategy. The H3K36me3 negative samples show substantial increase in the percent of reads mapping to Drosophila, which indicates a immunoprecipitation shift to the H3K36me3 positive Drosophila cells from the H3K36me3 negative human cells, and likely explains the differential findings of these two studies. This provides further support for the need to incorporate a normalization strategy when conducting sequencing experiments, especially when levels of the mark of interest are diminished.

This chapter also expanded on current knowledge of wildtype SETD2 by examining the coding exon preference of H3K36me3. The level of methylation did not correlate with exon length, but overall gene methylation did correlate with gene expression. This data enforces the link between H3K36me3 and transcription by suggesting gene expression is the key determinant of H3K36me3 levels. Additional work examining the role of nucleosome occupancy in H3K36me3 enrichment in coding exons will further explore this phenotype, as this has been suggested as a source of H3K36me3 enrichment in exons [63].

The data from SETD2Δ+tSETD2 samples suggest that the N-terminus is not necessary for faithful H3K36me3 placement. This is an intriguing finding, as the N-terminus region is not present in yeast Set2, but is present in many of the SETD2 homologs found in other organisms, including *P. troglodytes*, *C. lupus*, *B. taurus, M. musculus*, *D. rerio*, and *D. melanogaster*. This suggests that the N-terminus was evolutionarily selected, and therefore would be predicted to have a role in SETD2 function. Our data suggest that this role is not directly invovled in H3K36 trimethylation regulation.

One facet of SETD2 biology which has made study of this protein difficult is the low levels of expression for SETD2 in human cells. This was recently explored, and a region from amino acid 1104-1403 was shown to harbor a protein degradation signal which was MG132 dependent [165]. In addition, a region from amino acid 504-803 may provide a signal for protein stability. This suggests the N-terminus is required for regulating protein levels, but not for functional activity. tSETD2 begins at amino acid 1323 and thus lacks the latter region entirely, and contains less than 100 amino acids of the former region, and this may be why visualization of tSETD2 is possible while visualization of SETD2 is difficult. It is possible that the proteasome degradation sequence found from amino acid 1104-1403 is still present. Further detailed biochemical examination of this region will be required to determine the final protein sequence responsible.

Interestingly, tSETD2 often showed lower levels of H3K36me3 compared to 786O+EV. This could be due to the low levels of expression of tSETD2 compared to the other mutants. This may suggest that protein levels are very tightly regulated, and do contribute to over H3K36me3 levels. However, tSETD2 is expressed at levels most similar to wildtype SETD2. An additional argument against the low expression of tSETD2 contributing to low H3K36me3 signal is the high levels of the R2510H and T2457* mutant proteins do not fundamentally alter the H3K36me3 localization. It is also possible that the low levels of tSETD2 are due to a lower percentage of GFP positive cells. If a smaller percentage of cells are positive for GFP (and by proxy tSETD2), this would shift the normalization strategy and impact downstream analyses. tSETD2 did have a significant difference between input percent Drosophila and sample percent Drosophila, which would suggest a shift from human cells to Drosophila cells in this sample. This shift is the most likely source of decreased signal in these samples.

We have shown that mutation or deletion of the SRI domain disrupts the interaction between tSETD2 and RNAPII. Given this, the finding that SETD2Δ+R2510H and SETD2Δ+T2457* samples have very little of variation from 786O+EV was surprising. We had hypothesized that the interaction with RNAPII would be crucial to proper H3K36me3 placement. Instead, coding exons still showed the highest level of enrichment. Additionally, the number of coding exons which were significantly different between SRI altered tSETD2 and wildtype SETD2 was in the single digits when compared to 786O+EV. It is possible that there are differences due to SRI domain loss in regions of the genome which were not examined in this manuscript, however, the most prominent features of H3K36me3 are unaltered. This leaves the question of the role of the SRI domain mutation in ccRCC somewhat open.

Evidence has shown the SRI domain interacts with the hyperphosphorylated form for RNAPII [5]. Hyperphosphorylated RNAPII is phosphorylated on serine-2 and serine-5 of the C-terminal repeat domain, and is the elongating form of polymerase (reviewed in [166]). The SRI domain has been shown to have no affinity for RNAPII when it is only phosphorylated at serine-2 or serine-5 [5]. It is possible that SETD2 is interacting with RNAPII at an undetectable levels *in vitro* through a non-SRI mechanism. Were this the case, we would expect the interaction to be equivalent between hyperphosphorylated and unphosphorylated, and that the SRI altered forms of tSETD2 would have a skewed interaction for the unphosphorylated. As unphosphorylated RNAPII is found at promoters and early in the gene, the 5' bias of SETD2Δ+R2510H and SETD2Δ+T2457* would support this possibility. The HA-RNAPII co-immunoprecipitation in this study was done using an antibody which recognizes both the phosphorylated and unphosphorylated forms of RNAPII. Studies using antibodies which recognize particular phosphorylation states could help parse this interaction.

Overall, this data supports that of Chapter 2, showing total loss of H3K36me3 with TALEN inactivation of SETD2, as well as loss of catalytic activity by the R1625C mutant. Furthermore, it defines a separation of function for the SET and SRI domains of SETD2. Future work will explore the role of the SRI domain in additional processes to determine how a ccRCC-associated mutation in SETD2 that does not alter H3K36me3 placement contributes to cancer development.

4.4 Methods

4.4.1 Stable expression of tSETD2 and mutants

To generate stably expressing cells, 786O parental, 786O SETD2Δ, HKC parental and HKC SETD2Δ cells were transduced with empty vector (EV), tSETD2, R1625C, R2510H, or T2457* virus. After transduction, cells were sorted by GFP status into 96 well plates, 5 cells per well. After two weeks, the cells were checked for proliferation and GFP levels. Those wells which had proliferated and were 100% GFP positive were combined (3-5 wells) into a 24 well plate. This population was expanded and sorted again for the highest intensity GFP positive cells.

4.4.2 Flow Cytometry

Cells were grown on 10cm^2 plates to near confluency, isolated and pelleted. The pellets were washed with PBS, and stained with propidium iodide (PI) as a live-dead discriminator. Flow cytometry for PI and GFP was conducted on the Becton Dickinson LSRFortessa. Plots were generated using the Flowing Software version 2.5.1.

4.4.3 Protein extraction

Histones were extracted for histone western blots using an acid extraction protocol. Briefly, cells were pelleted and washed in cold PBS, then resuspended in TEB lysis buffer (PBS containing 0.5% Triton X 100 (v/v), 2mM phenylmethylsulfonyl fluoride (PMSF), 0.02% (w/v) NaN3). Cells were lysed by nutation for 10 minutes at 4°C. Lysate was pelleted, washed, and overnight acid extracted in 0.2N HCl. RIPA extracts were used for HA and GFP blots. RIPA (25mM Tris, 150mM NaCl, 0.1% SDS, 0.5% Na-deoxycholate, 1% Triton X/NP-40) was added to cold cell pellets, and incubated on ice 10 minutes. Lysates were centrifuged at high speed and supernatant was retained. For chromatin fractionation cell pellets were snap frozen in a dry-ice ethanol bath. Pellets thawed on ice, then were resuspended in CSK buffer (10mM Pipes, pH 7 with NaOH, 300mM sucrose, 100mM NaCl, 3mM $MgCl_2$, 0.1% Triton-X 100, 1x protease inhibitors, 1mM NaF). Sample was centrifuged, and supernatant retained as the soluble fraction. Pellet was washed in CSK, and resuspended in 1mL CSK as the chromatin fraction, which was used for immunoprecipitation.

4.4.4 Western blots

For histone western blots, 4-15% Mini-PROTEAN® TGX Stain-Free™ Protein Gels (Bio-Rad) were used. For HA and RNAPII blots, Any kD™ Mini-PROTEAN® TGX Stain-Free™ Protein Gels (Bio-Rad) were used. Antibodies used were anti-H3 (abcam ab10799), anti-H3K36me3 (abcam ab9050), anti-HA (Biolegend 901502), anti-RNA Pol II (Active Motif 39097), anti-tubulin (Sigma T9026), and anti-GFP (Cell Signaling Technology #2555). Secondary antibodies were LI-COR IRDye® 800CW Goat anti-Mouse, LI-COR IRDye® 680CW Goat anti-Mouse, LI-COR IRDye® 800CW Goat anti-Rabbit, LI-COR IRDye® 680CW Goat anti-Rabbit. Transfers were onto a nitrocellulose membrane, and were either 100V for 1 hour for histone blots or 30V for 12 hours for HA and RNAPII blots. 5% BSA in PBS was used for blocking, and antibodies were diluted in 5% BSA in PBS-T.

4.4.5 Co-immunoprecipitation

293T cells were transiently transfected using the TransIT®-LT1 Transfection Reagent with 10ug of empty vector, tSETD2, R2510H or T2457* plasmid. HKC stable expression cells were generated as described above. Cell pellets were resuspended in high salt IP buffer (50mM Tris pH 8, 300mM NaCl, 10% Glycerol, 1% NP-40, 20mM NaF, 10mM Sodium Pyrophosphate, 10mM Sodium Orthovanadate, 1x protease inhibitors), lysed by pipetting, and nutated at 4°C 30 minutes. The sample was spun and transferred to a new tube, where it was diluted 1:1 with no salt IP buffer (50mM Tris pH 8, 10% Glycerol, 1% NP-40, 20mM NaF, 10mM Sodium Pyrophosphate, 10mM Sodium Orthovanadate, 1x protease inhibitors), nutated at 4°C 30 minutes, then 2µg anti-HA (Biolegend 901502) was added and nutated overnight at 4°C. Washed Bio-Rad Protein G Surebeads (Bio-Rad 1614023) were added to the sample and nutated 2 hours at 4°C. The supernatant was discarded and the beads were wash 4x with mid-salt IP buffer (50mM Tris pH 8, 150mM NaCl 10% Glycerol, 1% NP-40, 20mM NaF, 10mM Sodium Pyrophosphate, 10mM Sodium Orthovanadate, 1x protease inhibitors). The sample was eluted by addition of 40uL of Laemmli buffer (63mM Tris-HCl, 10% Glycerol, 2% SDS, pH 6.8) and incubation at 70°C for 10 minutes. For chromatin fraction immunoprecipitation, fraction was generated as described above, and incubated overnight with 2µg anti-H3K36me3. Immunoprecipitation proceeded as above, with wash in CSK rather than mid-salt buffer.

4.4.6 ChIP-Rx

Cells used included 786O+EV, SETD2Δ+EV, SETD2Δ+tSETD2, SETD2Δ+R1625C, SETD2Δ+R2510H and SETD2Δ+T2457*. Two confluent 15cm^2 plates per cell type were fixed by addition of formaldehyde to a final concentration of 1%. Formaldehyde was inactivated with 125mM glycine, then cells were washed, collected by scrapping and pelleted. The pellets were resuspended in hypotonic homogenization buffer (10mM Tris pH 7.4, 15mM NaCl, 60mM KCl, 1mM EDTA, 0.1% NP-40, 5% sucrose, 1x protease inhibitors (Roche)) and dounced with 10 strokes. A sucrose pad (10mM Tris pH 7.4, 15mM NaCl, 60mM KCl, 10% sucrose, 1x protease inhibitors) was added to the bottom of samples, and samples were centrifuged. The supernatant was removed, and isolated nuclei were stored at -80.

Frozen nuclei were thawed on ice, 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 an average fragment size of 500 base pairs. Anti-H3K36me3 (abcam ab9050, Lot GR233723-2) antibody was added to Bio-Rad Protein A Surebeads (Bio-Rad 1614013), rotated for 10 minutes, washed, then 400uL of sonicated sample was added to the beads. After 1 hour of rotation at room temperature, the supernatant was discarded and the beads were wash 4x with PBS-T. The sample was eluted by addition of 40uL of Laemmli buffer and incubation at 70°C for 10 minutes. Elution was treated with RNAse-A and Proteinase

K, left at 65°C overnight, and then cleaned on Zymo ChIP DNA Clean & Concentrator columns (Zymo D5201).

4.4.7 Library preparation

Sequencing libraries for ChIP DNA all followed the standard Illumina sequencing protocol. 10ng of DNA was used as starting material. Libraries were generated for both the ChIP DNA and the DNA from the ChIP input. DNA was blunted, purified using Ampure XP beads, A-tailed, adapter ligated, purified using Ampure beads two additional times, before PCR amplification (15 cycles) and one additional purification. Final libraries were submitted as a pool to the UNC High Throughput Sequencing Facility (HTSF) for sequencing. ChIP-seq utilized 9 lanes of single-end 50bp reads.

4.4.8 Sequencing Pipeline

Data was obtained from UNC HTSF, then processed by the in-house sequencing pipeline. This pipeline processed the files to remove adapters using TagDust [145]. Paired-end sequences were synchronized to each other, then aligned to the genome using Bowtie [146]. The bowtie output files were convered to BAM files using samtools [147], then filtered for properly mapped reads. These files were used to generate downstream files for additional analyses.

Data was normalized to drosophila DNA content. The number of reads mapping to the drosophila genome and the number of reads mapping to the human genome was determine, which those ambiguously mapped to both removed. The normalization constant was determined as α=1/count of reads (in millions) aligning to the Drosophila genome, and is determined for each file. During wiggle file generation, the base count is multiplied by this normalization constant. Wiggle files have read extension to 500 base pair from the 50bp reads from the sequencer, as this was the average fragment length as determined by bioanalyzer tracings (data not shown).

4.4.9 Data analysis

Metagene analysis was conducted using CEAS [160]. CEAS generates the signal across each gene, and outputs an average value across a 3000 range. Genes which are smaller than 3000 will be expanded to fill this range, while genes larger than 300 will be compressed. The CEAS computed value was read into R [167], where the average and standard deviation for the 3 sample replicates was determine. The average value was plotted as a line, with shaded standard deviation added.

Signal enrichment was calculated using Zinba [149] and R. Genomic regions were downloaded from the UCSC Table browser [150]. Promoters were defined as 1kb upstream of the TSS and downloaded as such. Intergenic regions were generated by subtracting genes and promoters from the total genome. All other regions were defined through the UCSC download tool. Once files were downloaded, they were converted to Zinba format. Wiggle files for each sample and input were used to determine the signal at every base pair of the genomic region. The average signal per region was determined using an in-house perl script. The average signal per region was then read into R, and the sample signal was divided by the input singal, then this was averaged between replicates. The final signal was then log2 transformed, and plotted as boxplots using R. Zinba was also used to find signal at the TSS.

Density plots were also generated in R using the densCols function. densCols determines the density at each point. The ColorRampPalette function can then be used to assign a color to this density, with 256 color splits. This density can then be plotted as a scatterplot.

Principal component analysis was also completed in R using the prcomp function. Samples and inputs were divided into 500 base pair windows of average signal per window. A matrix of these files was generated, which was then used by the prcomp function, with center and scale set as true. Plots are of the first principal component versus the second.

Dendogram analysis was competed using the Benjamini-Hochberg method [163]. The average signal per region files described above were used as an input. Unpaired t-tests between samples were conducted for each region, and the false discovery rate was set at $\alpha<0.05$. The number of regions which reject the null hypothesis that there is no difference between samples was output, and used to generate a dendogram. The axis of this dendogram represents the number of regions.

CHAPTER 5: Conclusions and Discussion

5.1 Overall Summary

The focus of this work has been on SETD2 loss and mutation, specifically in the context of clear cell renal cell carcinoma. In Chapter 1, the overall role of SETD2 in the cell is discussed. SETD2 is a non-redundant histone H3 lysine 36 methyltransferase [1], as we have verified by ChIP-seq in Chapter 4. It has been shown to be involved in transcription [61,62,64], RNA processing [64,70–72], and DNA damage repair [83,84,86]. The role of SETD2 mutation in cancer is still largely unexplored. This work attempts to fill in the knowledge gap in ccRCC development.

We selected two ccRCC-associated mutations for initial study. The first chosen mutation, R1625C, is found in the SET domain. The second, R2510H, occurs in the SRI domain. We inactivated SETD2 in human cells using TAL effector nucleases and introduced a truncated but functional wildtype SETD2 (tSETD2), as well as the two mutants in the tSETD2 background into these cells. We found that SETD2 inactivation resulted in global loss of H3K36me3, which was not rescued by expression of the R1625C mutant. Expression of tSETD2 and R2510H restored H3K36me3 globally. We show that loss of catalytic activity in the R1625C mutant line was due to reduced substrate-binding capacity of this variant.

To further study ccRCC-associated mutations, we took advantage of the similarities between human SETD2 and the yeast ortholog of SETD2, Set2. Set2 has all of the functional domains found in SETD2, and is responsible for all levels of H3K36 methylation in yeast. We generated the ccRCC-associated mutations in Set2, as well as an additional H199L mutation. We found that the R2510H yeast equivalent mutation, K663L, did not alter H3K36me3 levels, while the R1625C equivalent, R195C, abolished H3K36me3 abilities of Set2 without altering H3K36me2 or H3K36me1. The H199L mutation resulted in loss of both H3K36me3 and H3K36me2. With this series of mutants, we were able to examine the effect of mutation on known Set2-associated phenotypes. We found that cryptic initiation, phleomycin sensitivity and 6-AU treatment resistance were all dependent on dimethylation rather than trimethylation. Based on this result, we examined similar features in human cells. We found that γH2A.X levels after DNA damage were higher in H3K36me3 deficient samples, but that this did not alter survival after DNA damage, in agreement with the yeast results.

We further explored the effects of SETD2 loss by examining SETD2Δ cells for changes in transcription and chromatin organization. Previous studies had shown widespread changes in chromatin accessibility in SETD2 mutant ccRCC tumors [17]. We hypothesized that these changes were due to alterations in nucleosome positioning, and chose to study this using MNase-seq. We did not observe widespread changes in nucleosome positioning at the TSS or internal exons, though we did show increased gene expression correlated with increased MNase signal at the TSS. In order to examine changes in chromatin accessibility more broadly, we conducted FAIRE-seq in the SETD2 wildtype and SETD2Δ cells. We found approximately 8000 regions of altered accessibility, with the majority showing increased accessibility in the SETD2Δ cells. This was in agreement with the results seen in tumors. We then examined the properties of the nucleosomes found in regions of increased accessibility, and found nucleosomes in these regions show decreased positional conservation, increased occupancy, but fewer overall nucleosomes. This suggests that there is wider variation in nucleosome positioning in regions of increased chromatin accessibility in SETD2Δ cells.

Studies of SETD2 mutant ccRCC tumors also showed widespread alterations in RNA processing. Based on this, we examined RNA in the SETD2Δ cells using RNA-seq. Surprisingly, in both the 786O and HKC cellular background we found very few genes that showed alterations in gene expression. Those that were significantly different often showed small changes. Additionally, the differential genes found in 786O did not overlap those found in HKC. Examining the SETD2Δ replicates showed substantial variation between replicates, which likely accounts for the low number of significantly different genes. Examining RNA splicing showed similar levels of intron retention between wildtype and knockout samples in both 786O and HKC samples. Overall, changes in RNA were found at low levels and were not consistent between replicates.

We next examined the effect of SETD2 mutation on H3K36me3 localization using a spike-in normalization strategy. SETD2 loss and the R1625C SET domain mutation result in complete loss of H3K36me3 genome-wide. We examined H3K36me3 localization in the 786O parental cells and showed signal increases along the length of the gene. Coding exons show the highest levels of enrichment for

H3K36me3, while non-genic regions show the lowest levels. H3K36me3 levels correlated with gene expression but not exon length. Clustering shows that H3K36me3 positive samples consistently cluster together, and have similar distances between them. SETD2Δ+tSETD2, SETD2Δ+R2510H, and SETD2Δ+T2457* samples all show a pattern similar to wildtype across genes and have high levels of H3K36me3 enrichment at coding exons.

Further examination of signal at the TSS showed that the SETD2Δ+T2457* had an increased level of signal early after the TSS. This was consistent regardless of the gene expression level. Examining specific loci showed that there was a shift in H3K36me3 levels towards the TSS. Interestingly, this shift was also seen in the SETD2Δ+R2510H when specific genes were examined. This suggests that the increase in signal near the TSS is due to lost interaction with the hyperphosphorylated RNAPII.

This work has expanded on the current understanding of SETD2 biology. The inactivation of catalytic activity by the R1625C mutation supports the role of H3K6me3 loss in ccRCC development. The fact that the SRI domain does not alter H3K36me3 levels or localization is a novel finding, and brings into question the relationship between SETD2 and RNAPII. This separation of function between domains will be of key importance when designing therapeutics which target SETD2 mutation in ccRCC. Future studies will expand on this novel finding by exploring the relationship between the SRI domain and ccRCC development.

5.2 Transcription and Chromatin

One the surprises of this work was the relatively low amount of variation observed between SETD2 wildtype and SETD2Δ cells in chromatin and transcription assays. Given the substantial amount of published work that showed a role for H3K36me3 in transcription and RNA processing, we predicted that loss of SETD2 would have a large effect on RNA abundance and splicing. Ultimately, this is not what was observed. What was notable was that there were large numbers of genes which showed high levels of variation, but these genes were not consistent between samples. This opens an intriguing possibility. As SETD2 and H3K36me3 have been implicated in DNA damage repair, it is possible that the cells used for these studies have diverged during growth and passage. The longer the cells were grown, the further the SETD2Δ replicates diverged from each other. As a result, their gene expression profiles would likely change as well. The process by which the SETD2Δ replicates were generated started at a single cells stage, which then grew into a population of cells over the course of many weeks. Within this population it is also possible that subpopulations may have emerged if DNA damage occurred and was poorly repaired.

The work from ccRCC tumors would support this as a possibility. The ccRCC TCGA data analysis did not identify a consistent gene expression signature specific to SETD2 mutant tumors [18]. One way to test for this would be to repeat RNA sequencing with a technical replicate rather than biological replicates. If the number of statistically significant genes drastically increases and there is still little overlap between different SETD2Δ samples, there is likely not a consistent pattern of gene expression related to SETD2 loss. This work provides support for this hypothesis. Future studies can expand on this finding by showing consistency between technical replicates, as well as showing changes in gene expression which occur over time. Inducible loss of SETD2 would allow examination of RNA immediately after SETD2 loss, rather than after several cell divisions. This would allow comparison between acute and chronic SETD2 loss, which would provide more insight into early ccRCC development after SETD2 mutation.

We also examined the effect of SETD2 loss on chromatin organization. Again, we were surprised that changes in nucleosome positioning were not more widespread. Evidence from SETD2 mutant tumors suggested that altered chromatin organization at exons was due to altered nucleosome positioning. We did not observe changes in nucleosome positioning at internal exons. We did, however, see changes in nucleosome properties at regions with altered chromatin accessibility. Recent studies have shown that changes in chromatin accessibility can occur without changes in nucleosome positioning [142]. Studies in yeast have shown that H3K36me2 recruits the Rpd3s histone deacetylase complex to genes, resulting in decreased acetylation [112]. Additionally, H3K36me3 in yeast recruits the NuA3b acetyltransferase complex through the PWWP domain of Pdp3 [168]. This suggests that loss of H3K36me3 could result in decreased acetylation. It is also possible H3K36me3 may be recruiting additional chromatin modifiers to genes, altering the chromatin landscape. To test this, the acetylation state of individual and neighboring nucleosomes could be studied. One method to examine modifications on individual nucleosomes is Combinatorial-iChIP. This process identifies modifications which are found on the same nucleosomes using a dual barcoded ChIP re-ChIP method [169]. This would allow examination of various acetyl states on H3K36me3 marked nucleosomes. Marks which are common to these nucleosomes can then be studied in the H3K36me3 negative cell lines for changes using ChIP-Rx. Additionally, we could develop a process to study neighboring nucleosomes. This process might use MNase to generate dinucleosomes, which are then immunoprecipitated for the mark of interest (i.e. H3K36me3). The dinucleosomes could then be cleaved to mononucleosomes, which are then queried for various modifications. This would enable examination of the ‘next’ nucleosome’s modification state. In general, understanding the precise state of a region of chromatin will provide further insight into the specific processes which are affected by the loss of a particular mark. For SETD2 biology, this would provide a comprehensive understanding of how chromatin is affected by SETD2 loss in ccRCC.

5.3 The role of the SRI domain

The most surprising finding of this work was the relatively normal genomic targeting of H3K36me3 in SETD2Δ+R2510H and SETD2Δ+T2457* cells. This contradicts the traditional view of SETD2 interacting with RNAPII to mark nucleosomes within a gene. We show that both SRI domain mutations disrupt the interaction between SETD2 and RNAPII, so this finding indicates that this interaction is not necessary for normal H3K36me3 placement. If this is true, the obvious question then becomes “What is the role of the SRI domain?”

The focus of this work has been on the function of SETD2 as a histone methyltransferase. This is a logical choice, as until recently H3K36 was the only known target of SETD2. However, recent studies have questioned this, and at least one new target has been identified. A recent study (which we were involved in as collaborators) showed that SETD2 also targets tubulin for methylation [109]. Loss of this tubulin mark leads to mitotic and cytokinetic errors, which likely contributes to genome instability in ccRCC. The study examined the effect of ccRCC-associated mutations on tubulin K40me3 levels. R1625C expression failed to restore this mark in SETD2Δ 786O cells, which would be expected as it is catalytically inactive. Surprisingly, the expression of the R2510H mutant in SETD2Δ also failed to restore tubulin K40me3. This suggests that the SRI domain is required for tubulin methylation, and our data suggest it is not required for histone methylation. Further exploration of the tubulin methylation will be required to verify the role of the SRI domain. If the SRI domain is required, we would observe no tubulin K40me3 with expression of the T2457* SRI deletion. In addition, expression of tSETD2 should rescue mitotic defects, while expression of R2510H or T2457* will not. This work is currently being examined in our lab and that of our collaborators.

It is possible that the SRI domain only stabilizes the interaction between the RNAPII elongating complex and SETD2 through the interaction with the hyperphosphorylated CTD. The elongating complex consists of many proteins, including cleavage and polyadenylation factors (reviewed in [166]). Without the stabilization provided by the SRI domain, T2457* may still be capable of interaction with other members of this complex. A comprehensive analysis of the proteins that interact with SETD2 would provide further insight into how SRI altered tSETD2 is able to mark genic regions in a similar manner to wildtype, despite lost interaction with RNAPII. This could be completed by mass spectrometry or yeast two-hybrid screening for protein partners. As more methylation targets and interacting partners for SETD2 are identified, it will be crucial to examine the role of the SRI domain in these functions.

5.4 SETD2 mutation and therapeutics

ccRCC is one of the ten most common cancers in both men and women in the United States [134]. SETD2 is mutation in ccRCC has been associated with decreased overall survival and is a univariate predictor of survival [103]. SETD2 mutant tumors are also more likely to present with stage III or higher disease [104]. The state at which a tumor is found is a key determinant of survival. Five year survival for ccRCC in the United States from 2005-2011 was 92% for localized disease, 65% for regional disease, and 12% for metastatic disease (reviewed in [170]). The need for better therapeutics in metastatic disease is obvious. Recent studies have explored SETD2 loss as a therapeutic target for ccRCC. The Genomics of Drug Sensitivity in Cancer database [171] lists 4 compounds which selectively affect *SETD2*-/- cell lines (reviewed in [172]). Of these, two are PI3Kβ inhibitors. One study examined the effects of PI3Kβ inhibition in ccRCC cells and found that TGX221 selectively inhibited ccRCC cells with both VHL and SETD2 mutations [173]. A different study found that WEE1 inhibition in H3K36me3-deficient cells results in RRM2 reduction, critical dNTP depletion, S-phase arrest, and apoptosis [95].

One key feature of both these studies is they examine the effect of the compound of interest in the context SETD2 and H3K36me3 loss. The work in this dissertation however suggests the H3K36me3 loss and SETD2 mutation are not equivalent. Targeting SETD2 loss is effective if the ccRCC tumor has mutations which are early inactivating or occur in the SET domain. We have determined, however, that SRI domain mutation is fundamentally different from SETD2 loss. There is little to no effect on H3K36me3 in SRI mutant cells, thus limiting the therapeutic focus in ccRCC to H3K36me3 loss will not help develop treatments for SRI mutant tumors. Treating SETD2 mutant tumors will thus have to encompass two tasks: directing therapies at H3K36me3 loss and targeting SRI domain alteration.

Therapeutically targeting SRI domain mutations is an unexplored field. As stated above, the SRI domain has been implicated in tubulin methylation. It may be possible to treat SRI domain mutants by inhibiting mitosis. There are several chemotherapeutics used in clinic to target the mitotic checkpoint, including taxanes and vinca alkaloids [174]. A recent study of the Aurora kinase inhibitor VX680/MK-0457 in ccRCC cells showed inhibition of cell growth *in vitro* [175]. Aurora kinases are regulators of mitosis, and have previously been implicated in tumorigenesis. They are also upregulated in ccRCC compared to normal kidney. It is possible that this inhibitor may specifically aid in the treatment of SRI mutant ccRCC, and may be less of effective in ccRCC with other types of SETD2 mutation.

5.5 In conclusion

SETD2 mutation is a common event in ccRCC that is associated with worse prognosis. Understanding of these mutations is critical to developing therapeutics to treat this disease. Mutations in SETD2 which are early inactivating or interrupt the catalytic SET domain result in complete loss of H3K36me3, while those in the SRI domain have little effect on this mark. These results show the importance of studying specific cancer-associated mutations rather than focusing on protein loss. Had we limited our work to specifically studying the effect of SETD2 loss, we would have concluded that the role of SETD2 mutation in ccRCC development was dependent on H3K36me3 loss. By conducting a domain-specific analysis, the novel finding that the SRI domain is not required H3K36me3 placement was identified. Future work will further explore the role of the SRI domain in cancer development, by investigating the effect on tubulin methylation, identifying additional interacting partners, and examining therapeutic opportunities specific to this mutation.

APPENDIX: A Tale of Two Cancers - Complete genetic analysis of Chromophobe Renal Cell Carcinoma contrasts with Clear Cell Renal Cell Carcinoma[3]

The Cancer Genome Atlas projects in rare tumor types offer unique insights into mechanisms of tumorigenesis. The first such project studied Chromophobe Renal Cell Carcinoma, ChRCC [176]. ChRCC represents 5% of renal cell carcinoma cases [177], and is associated with a striking aneuploidy pattern [178]. Our analysis of 66 cases of non-hereditary ChRCC included whole-exome DNA sequencing for all cases, along with whole-genome DNA sequencing in 50 of these cases, and 61 cases with mtDNA sequencing by long range PCR. All cases also were studied for copy number analysis, mRNA and miRNA sequencing, and CpG DNA methylation.

Our analyses verified large-scale chromosomal loss that has previously been described, with 86% of cases showing loss of one copy of the entire chromosome, for most or all of chromosomes 1, 2, 6, 10, 13, and 17. Between 12% and 58% of cases also showed entire chromosomal loss for chromosomes 3, 5, 8, 9, and 21. Chromophobe has also been defined with histological classifications, classic and eosinophilic. Of the 47 tumors with classic histology, all showed the characteristic chromosomal loss, while only 10 of the 19 eosinophilic cases showed these losses. Additionally, eosinophilic cases showed no chromosomal loss, suggesting a degree of genomic heterogeneity as a distinguishing characteristic of ChRCC histology classification.

Whole exome sequencing showed relatively low numbers of somatic mutations, with only 2 reaching a threshold of being called frequently mutated genes (occurring in >5% of cases). *TP53* mutations were identified in 21 cases (32%), and *PTEN* was mutated in 6 cases (9%). With only two common mutations, our analyses turned to methylation changes, mitochondrial DNA, and structural changes.

By comparing methylation patterns between the most common subtype of RCC, clear cell renal cell carcinoma (ccRCC), and ChRCC, we were able to identify distinct differences between these two subtypes, suggesting a potential difference in cell of origin. To further examine this possibility, we compared RNA seq gene expression data from both diseases to an external gene expression data set of normal kidney which had been microdissected from various regions of the nephron [179]. ChRCC mRNA expression highly

[3] This research has been published in Molecular & Cellular Oncology. Fahey CC, Rathmell WK. A tale of two cancers: Complete genetic analysis of chromophobe renal cell carcinoma contrasts with clear cell renal cell carcinoma. *Mol Cell Oncol.* 2015;2: e979686. doi:10.4161/23723556.2014.979686

correlated with expression profiles from distal regions of the nephron, while ccRCC mRNA expression correlated with expression profiles from the proximal nephron. These results suggest that molecular differences between these subtypes may be defined and influenced by distinct cells of origin.

Bioenergetic features were prominent in both ccRCC and ChRCC, but in highly divergent patterns. In ChRCC, nearly all genes involved in the Krebs cycle showed increased expression when compared to normal kidney, as did at least one gene for each complex involved in the electron transport chain (ETC). There were also increased mitochondrial genome numbers in ChRCC. Interestingly, mitochondrial related genes are strongly repressed in ccRCC, highlighting another important distinction between these two subtypes [18]. Additionally, these differences suggest alternate strategies to support tumor growth rather than minimizing reliance on oxidative phosphorylation.

Further mitochondrial analysis revealed a high frequency of somatic mutations in mitochondrial genes at various levels of heteroplasmy. 12 tumors had nonsilent mutations, including many frameshift mutations. Mitochondrial gene mutations have been described previously [180], but this analysis confirms both the high frequency of events, and a close link of these mutations to the eosinophilic subgroup. Generally, these mutations were predicted to be inactivating. Specific mutations were identified in *MT-ND5*, an essential member of ETC complex I, including one previously identified cancer-associated mutation [181]. Intriguingly, mutations in complex I were not correlated with alterations in expression patterns associated with loss of oxidative phosphorylation. This suggests an alternate role for complex I alteration in cancer metabolism, or a compensatory mechanism of increased gene expression, that does not reflect the same change in metabolic activity. Ultimately, measurements of substrate utilization and dynamic metabolic processing will be required to resolve the true metabolic state of these tumors.

The whole genome analysis identified two novel findings in ChRCC cases. A subset of samples displayed kataegis, the occurrence of highly localized substitution mutations (C>T or C>G) [182,183], which were localized with regions of genomic rearrangements. Among the samples with a strong kataegis pattern, TERT was identified as a differentially expressed gene. To further understand the role of TERT in ChRCC, we examined the sequence of the TERT promoter, and identified 3 cases with the previously described C228T mutation [184]. In addition, several cases displayed novel genomic rearrangements involving the TERT promoter, which were associated with the highest level of TERT expression. These variants were observed with a very high allelic frequency, suggesting TERT-associated rearrangements are early, potentially driver events. No equivalent findings have been reported in ccRCC.

As the first comprehensive molecular analysis of ChRCC, our studies identified several unique characteristics of this tumor type, which further reinforce its position as a distinct tumor entity from ccRCC. The combination of methylation pattern differences and expression correlations suggest a distal nephron cell of origin, while indicating proximal nephron origin of ccRCC. The finding of mtDNA alterations raise intriguing questions about the divergent strategies of kidney cells to induce altered metabolism in cancer, as our data suggested a complex metabolic phenotype due to more than loss of oxidative phosphorylation.

Finally, the finding of genomic rearrangements of the TERT promoter suggests alteration of TERT expression as a potential driver of this cancer (Figure A.1). These findings were only possible due to the integrated analysis of several data types, and the extension of sequencing beyond the confines of the exome. The comparison to ccRCC further reinforces the distinct origins and tumorigenic features of these cancers, which has important implications for future therapeutic innovations.

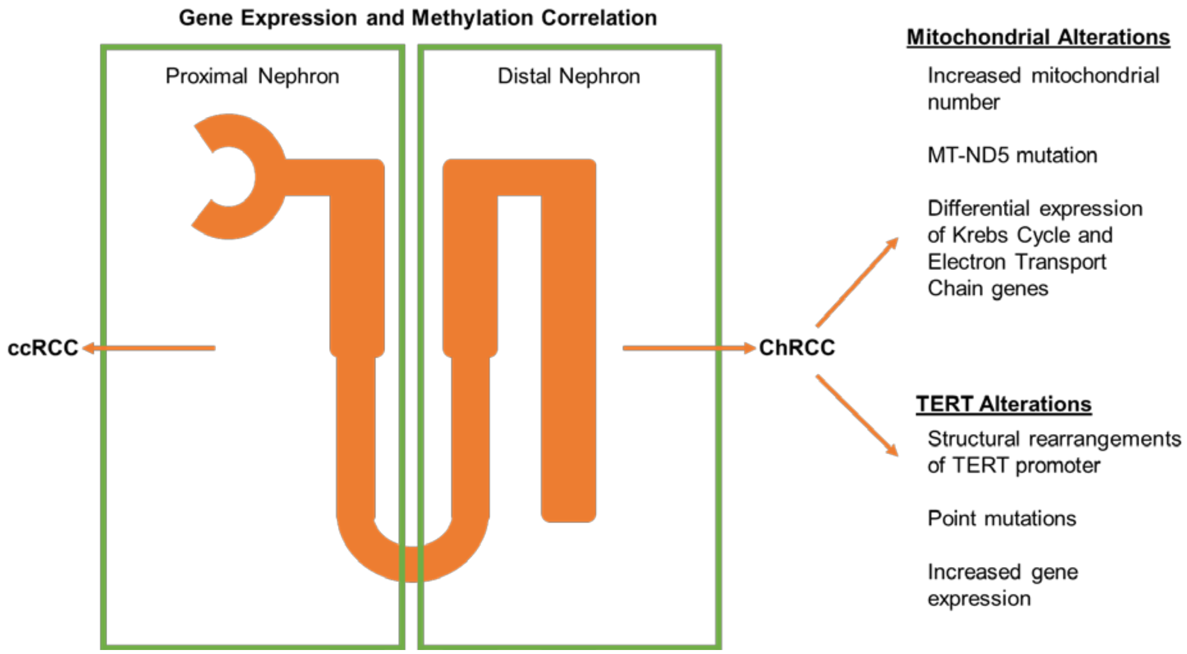

Gene Expression and Methylation Correlation
Proximal Nephron
Distal Nephron
ccRCC
ChRCC
Mitochondrial Alterations
Increased mitochondrial number
MT-ND5 mutation
Differential expression of Krebs Cycle and Electron Transport Chain genes
TERT Alterations
Structural rearrangements of TERT promoter
Point mutations
Increased gene expression

Figure A.0.1: Integrated Analysis identifies key characteristics of Chromphobe Renal Cell Carcinoma (ChRCC). ChRCC and clear cell renal cell carcinoma (ccRCC) originate from different regions of the kidney nephron. ChRCC is characterized by mitochondrial and TERT alterations.

BIBLIOGRAPHY

1. Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. 2008;27: 406–420.

2. Yuan W, Xie J, Long C, Erdjument-Bromage H, Ding X, Zheng Y, et al. Heterogeneous nuclear ribonucleoprotein L is a subunit of human KMT3a/set2 complex required for H3 Lys-36 trimethylation activity in vivo. J Biol Chem. 2009;284: 15701–15707.

3. Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Biol. 2012;13: 115–126.

4. Strahl BD, Grant PA, Briggs SD, Sun Z-W, Bone JR, Caldwell JA, et al. Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. Mol Cell Biol. 2002;22: 1298–1306.

5. Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P. 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. 2005;102: 17636–17641.

6. Sun X-J, Wei J, Wu X-Y, Hu M, Wang L, Wang H-H, et al. Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. J Biol Chem. 2005;280: 35261–35271.

7. Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH, et al. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. 2003;17: 654–663.

8. Li B, Howe L, Anderson S, Yates JR, Workman JL. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem. 2003;278: 8897–8903.

9. Gao Y-G, Yang H, Zhao J, Jiang Y-J, Hu H-Y. Autoinhibitory structure of the WW domain of HYPB/SETD2 regulates its interaction with the proline-rich region of huntingtin. Structure. 2014;22: 378–86.

10. Lu PJ, Zhou XZ, Shen M, Lu KP. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. Science. 1999;283: 1325–1328.

11. Gao J, Aksoy B, Dogrusoz U, Dresdner G. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science. 2013;6: 1–20.

12. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2: 401–404.

13. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010;463: 360–363.

14. Duns G, Hofstra RMW, Sietzema JG, Hollema H, van Duivenbode I, Kuik A, et al. Targeted exome sequencing in clear cell renal cell carcinoma tumors suggests aberrant chromatin regulation as a crucial step in ccRCC development. Hum Mutat. 2012;33: 1059–1062.

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

16. Stolle C, Glenn G, Zbar B, Humphrey JS, Choyke P, Walther M, et al. Improved detection of germline mutations in the von Hippel-Lindau disease tumor suppressor gene. Hum Mutat. 1998;12: 417–23.

17. Simon JM, Hacker KE, Singh D, Brannon AR, Parker JS, Weiser M, et al. Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res. 2014;24: 241–50.

18. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499: 43–9.

19. Gossage L, Murtaza M, Slatter AF, Lichtenstein CP, Warren A, Haynes B, et al. Clinical and pathological impact of VHL, PBRM1, BAP1, SETD2, KDM6A, and JARID1c in clear cell renal cell carcinoma. Genes Chromosomes Cancer. 2014;53: 38–51.

20. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366: 883–92.

21. Ho TH, Park IY, Zhao H, Tong P, Champion MD, Yan H, et al. High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. Oncogene. 2015; 1–10.

22. Fontebasso AM, Schwartzentruber J, Khuong-Quang D-A, Liu X-Y, Sturm D, Korshunov A, et al. Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. Acta Neuropathol. 2013;125: 659–69.

23. Jones S, Stransky N, McCord CL, Cerami E, Lagowski J, Kelly D, et al. Genomic analyses of gynaecologic carcinosarcomas reveal frequent mutations in chromatin remodelling genes. Nat Commun. 2014;5: 5006.

24. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497: 67–73.

25. Mar BG, Bullinger LB, McLean KM, Grauman P V., Harris MH, Stevenson K, et al. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. Nat Commun. 2014;5: 3469.

26. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012;481: 157–63.

27. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014;507: 315–22.

28. Van Allen EM, Mouw KW, Kim P, Iyer G, Wagle N, Al-Ahmadie H, et al. Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. Cancer Discov. 2014;4: 1140–1153.

29. Shain AH, Garrido M, Botton T, Talevich E, Yeh I, Sanborn JZ, et al. Exome sequencing of desmoplastic melanoma identifies recurrent NFKBIE promoter mutations and diverse activating mutations in the MAPK pathway. Nat Genet. 2015;47: 1194–1199.

30. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, et al. Melanoma genome sequencing reveals frequent PREX2 mutations. Nature. 2012;485: 502–506.

31. Hodis E, Watson IR, Kryukov G V., Arold ST, Imielinski M, Theurillat JP, et al. A landscape of driver mutations in melanoma. Cell. 2012;150: 251–263.

32. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511: 543–50.

33. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012;150: 1107–1120.

34. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, et al. Recurrent R-spondin fusions in colon cancer. Nature. 2012;488: 660–664.

35. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487: 330–7.

36. Witkiewicz AK, McMillan E a., Balaji U, Baek G, Lin W-C, Mansour J, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nat Commun. 2015;6: 6744.

37. Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513: 202–9.

38. The Cancer Genome Atlas Research Network. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N Engl J Med. 2016;374: 135–45.

39. Li YY, Hanna GJ, Laga AC, Haddad RI, Lorch JH, Hammerman PS. Genomic analysis of metastatic cutaneous squamous cell carcinoma. Clin Cancer Res. 2015;21: 1447–1456.

40. Wu G, Diaz AK, Paugh BS, Rankin SL, Ju B, Li Y, et al. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. Nat Genet. 2014;46: 444–450.

41. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482: 226–231.

42. Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfort J, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet. 2012;44: 251–253.

43. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DTW, Konermann C, et al. Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. Cancer Cell. 2012;22: 425–437.

44. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature. 2009;462: 739–44.

45. Chowdhury R, Yeoh KK, Tian Y, Hillringhaus L, Bagg EA, Rose NR, et al. The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. EMBO Rep. 2011;12: 463–9.

46. Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell. 2011;19: 17–30.

47. Lu C, Ward P, Kapoor G, Rohle D. IDH mutation impairs histone demethylation and results in a block to cell differentiation. Nature. 2012;483: 474–478.

48. Zhu X, He F, Zeng H, Ling S, Chen A, Wang Y, et al. Identification of functional cooperative mutations of SETD2 in human acute leukemia. Nat Genet. 2014;46: 287–293.

49. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368: 2059–74.

50. Milne TA, Kim J, Wang GG, Stadler SC, Basrur V, Whitcomb SJ, et al. Multiple Interactions Recruit MLL1 and MLL1 Fusion Proteins to the HOXA9 Locus in Leukemogenesis. Mol Cell. 2010;38: 853–863.

51. Yokoyama A, Lin M, Naresh A, Kitabayashi I, Cleary ML. A Higher-Order Complex Containing AF4 and ENL Family Proteins with P-TEFb Facilitates Oncogenic and Physiologic MLL-Dependent Transcription. Cancer Cell. 2010;17: 198–212.

52. Lee JJ, Sholl LM, Lindeman NI, Granter SR, Laga AC, Shivdasani P, et al. Targeted next-generation sequencing reveals high frequency of mutations in epigenetic regulators across treatment-naïve patient melanomas. Clin Epigenetics. 2015;7: 59.

53. Parker H, Rose-Zerilli M, Larrayoz M, Clifford R, Edelmann J, Blakemore S, et al. Genomic disruption of the histone methyltransferase SETD2 in chronic lymphocytic leukaemia. Leukemia. 2016; 1–8.

54. Huang KK, McPherson JR, Tay ST, Das K, Tan IB, Ng CCY, et al. SETD2 histone modifier loss in aggressive GI stromal tumours. Gut. 2015;0: 1–13.

55. Tan J, Ong CK, Lim WK, Ng CCY, Thike AA, Ng LM, et al. Genomic landscapes of breast fibroepithelial tumors. Nat Genet. 2015;47: 1341–1345.

56. Liu S-Y, Joseph NM, Ravindranathan A, Stohr BA, Greenland NY, Vohra P, et al. Genomic profiling of malignant phyllodes tumors reveals aberrations in FGFR1 and PI-3 kinase/RAS signaling pathways and provides insights into intratumoral heterogeneity. Mod Pathol. 2016; 1–16.

57. Xiang W, He J, Huang C, Chen L, Tao D, Wu X, et al. miR-106b-5p targets tumor suppressor gene SETD2 to inactive its function in clear cell renal cell carcinoma. Oncotarget. 2015;6: 4066–4079.

58. Behjati S, Tarpey PS, Presneau N, Scheipl S, Pillay N, Van Loo P, et al. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nat Genet. 2013;45: 1479–82.

59. Lewis PW, Müller MM, Koletsky MS, Cordero F, Lin S, Banaszynski L a, et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science. 2013;340: 857–61.

60. Fang D, Gan H, Lee J-H, Han J, Wang Z, Riester SM, et al. The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science. 2016;352: 1344–8.

61. Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, et al. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Mol Cell Biol. 2003;23: 4207–18.

62. Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem. 2005;280: 17732–17736.

63. Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. Nat Struct Mol Biol. 2009;16: 990–5.

64. de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, et al. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nat Struct Mol Biol. 2011;18: 977–983.

65. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell. 2005;123: 581–592.

66. Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell. 2005;123: 593–605.

67. Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, Kumar S, et al. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature. 2012;489: 452–455.

68. Lickwar CR, Rao B, Shabalin AA, Nobel AB, Strahl BD, Lieb JD. The set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency. PLoS One. 2009;4.

69. Li B, Gogol M, Carey M, Pattenden SG, Seidel C, Workman JL. Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. Genes Dev. 2007;21: 1422–1430.

70. Kolasinska-zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet. 2009;41: 376–381.

71. Luco, Reini F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., Misteli T. Regulation of Alternative Splicing by Histone Modifications. Science. 2010;327: 996–1000.

72. Kim S, Kim H, Fong N, Erickson B, Bentley DL. Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proc Natl Acad Sci U S A. 2011;108: 13564–9.

73. Yoh SM, Lucas JS, Jones KA. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev. 2008;22: 3422–3434.

74. Kanu N, Grönroos E, Martinez P, Burrell R a, Yi Goh X, Bartkova J, et al. SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. Oncogene. 2015; 1–10.

75. Grosso AR, Leite AP, Carvalho S, Matos MR, Martins FB, Vítor AC, et al. Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. Elife. 2015;4: 1–16.

76. Singh DP, Kimura A, Chylack LT, Shinohara T. Lens epithelium-derived growth factor (LEDGF/p75) and p52 are derived from a single gene by alternative splicing. Gene. 2000;242: 265–273.

77. Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS Genet. 2012;8.

78. Guo R, Zheng L, Park JW, Lv R, Chen H, Jiao F, et al. BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. Mol Cell. 2014;56: 298–310.

79. Zhang P, Du J, Sun B, Dong X, Xu G, Zhou J, et al. Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Res. 2006;34: 6621–6628.

80. Xie L, Pelz C, Wang W, Bashar A, Varlamova O, Shadle S, et al. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. EMBO J. 2011;30: 1473–84.

81. Carvalho S, Raposo AC, Martins FB, Grosso AR, Sridhara SC, Rino J, et al. Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. Nucleic Acids Res. 2013;41: 2881–2893.

82. Jung H, Lee D, Lee J, Park D, Kim YJ, Park W-Y, et al. Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat Genet. 2015;47: 1242–1248.

83. Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, et al. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat Struct Mol Biol. 2014;21: 366–74.

84. Carvalho S, Vítor AACA, Sridhara SCS, Martins FB, Raposo AC, Desterro JMP, et al. SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. Elife. 2014;3: e02482.

85. Xie P, Tian C, An L, Nie J, Lu K, Xing G, et al. Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. Cell Signal. 2008;20: 1671–1678.

86. Pfister SX, Ahrabi S, Zalmas LP, Sarkar S, Aymard F, Bachrati CZ, et al. SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. Cell Rep. 2014;7: 2006–2018.

87. Daugaard M, Baude A, Fugger K, Povlsen LK, Beck H, Sørensen CS, et al. LEDGF (p75) promotes DNA-end resection and homologous recombination. Nat Struct Mol Biol. 2012;19: 803–10.

88. Symington LS. Mechanism and regulation of DNA end resection in eukaryotes. Crit Rev Biochem Mol Biol. 2010;51: 195–212.

89. Li F, Mao G, Tong D, Huang J, Gu L, Yang W, et al. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. Cell. 2013;153: 590–600.

90. Awwad SW, Ayoub N. Overexpression of KDM4 lysine demethylases disrupts the integrity of the DNA mismatch repair pathway. Biol Open. 2015;4: 498–504.

91. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010;285: 26114–26120.

92. Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature. 2015;520: 243–7.

93. Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP. Relationship between gene body DNA methylation and intragenic H3K9ME3 and H3K36ME3 chromatin marks. PLoS One. 2011;6.

94. Tiedemann RL, Hlady RA, Hanavan PD, Lake DF, Tibes R, Lee J-H, et al. Dynamic reprogramming of DNA methylation in SETD2-deregulated renal cell carcinoma. Oncotarget. 2016;7: 1927–46.

95. Pfister SX, Markkanen E, Jiang Y, Sarkar S, Woodcock M, Orlando G, et al. Inhibiting WEE1 Selectively Kills Histone H3K36me3-Deficient Cancers by dNTP Starvation. Cancer Cell. 2015;28: 557–568.

96. Zhang Y, Xie S, Zhou Y, Xie Y, Liu P, Sun M, et al. H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm. Cell Rep. 2014;8: 1989–2002.

97. Lu C, Jain SU, Hoelper D, Bechet D, Molden RC, Ran L, et al. Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science. 2016;352: 844–9.

98. Li J, Kluiver J, Osinga J, Westers H, van Werkhoven MB, Seelen MA, et al. Functional Studies on Primary Tubular Epithelial Cells Indicate a Tumor Suppressor Role of SETD2 in Clear Cell Renal Cell Carcinoma. Neoplasia. 2016;18: 339–46.

99. Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, et al. Haploinsufficiency of NSD1 causes Sotos syndrome. Nat Genet. 2002;30: 365–366.

100. Tlemsani C, Luscan A, Leulliot N, Bieth E, Afenjar A, Baujat G, et al. SETD2 and DNMT3A screen in the Sotos-like syndrome French cohort. J Med Genet. 2016;36: 1–9.

101. Luscan A, Laurendeau I, Malan V, Francannet C, Odent S, Giuliano F, et al. Mutations in SETD2 cause a novel overgrowth condition. J Med Genet. 2014;51: 512–517.

102. Lapunzina P, Cohen MM. Risk of tumorigenesis in overgrowth syndromes: A comprehensive review. Am J Med Genet - Semin Med Genet. 2005;137 C: 53–71.

103. Hakimi AA, Ostrovnaya I, Reva B, Schultz N, Chen YB, Gonen M, et al. Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: A report by MSKCC and the KIRC TCGA research network. Clin Cancer Res. 2013;19: 3259–3267.

104. Hakimi AA, Chen YB, Wren J, Gonen M, Abdel-Wahab O, Heguy A, et al. Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. Eur Urol. 2013;63: 848–854.

105. Wang J, Liu L, Qu Y, Xi W, Xia Y, Bai Q, et al. Prognostic value of SETD2 expression in patients with metastatic renal cell carcinoma treated with tyrosine kinase inhibitors. J Urol. 2016;

106. Newbold RF, Mokbel K. Evidence for a tumour suppressor function of SETD2 in human breast cancer: A new hypothesis. Anticancer Res. 2010;30: 3309–3311.

107. Al Sarakbi W, Sasi W, Jiang WG, Roberts T, Newbold RF, Mokbel K. The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters. BMC Cancer. 2009;9: 290.

108. Feng C, Ding G, Jiang H, Ding Q, Wen H. Loss of MLH1 confers resistance to PI3Kβ inhibitors in renal clear cell carcinoma with SETD2 mutation. Tumor Biol. 2015;36: 3457–3464.

109. Park IY, Powell RT, Tripathi DN, Dere R, Ho TH, Blasius TL, et al. Dual Chromatin and Cytoskeletal Remodeling by SETD2. Cell. 2016;166: 950–962.

110. Duns G, van den Berg E, van Duivenbode I, Osinga J, Hollema H, Hofstra RMW, et al. Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. Cancer Res. 2010;70: 4287–4291.

111. Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. Mol Cell Biol. 2005;25: 3305–3316.

112. Li B, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, et al. Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s Histone deacetylase complex and to repress spurious transcription. J Biol Chem. 2009;284: 7970–7976.

113. Quan TK, Hartzog GA. Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription. Genetics. 2010;184: 321–334.

114. Fnu S, Williamson EA, De Haro LP, Brenneman M, Wray J, Shaheen M, et al. Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proc Natl Acad Sci U S A. 2011;108: 540–5.

115. Sorenson MR, Jha DK, Ucles S a, Flood DM, Strahl BD, Stevens SW, et al. Histone H3K36 methylation regulates pre-mRNA splicing in Saccharomyces cerevisiae. RNA Biol. 2016;13: 412–26.

116. Wen H, Li Y, Xi Y, Jiang S, Stratton S, Peng D, et al. ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. Nature. 2014;508: 1–18.

117. Zheng W, Ibáñez G, Wu H, Blum G, Zeng H, Dong A, et al. Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. J Am Chem Soc. 2012;134: 18004–18014.

118. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015;12: 7–8.

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

120. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9: 40.

121. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 2015;43: W174–W181.

122. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet. 2008;Chapter 10: Unit 10.11.

123. Jha DK, Strahl BD. An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair. Nat Commun. 2014;5: 3965.

124. Vojnic E, Simon B, Strahl BD, Sattler M, Cramer P. Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. J Biol Chem. 2006;281: 13–15.

125. Pai C-C, Deegan RS, Subramanian L, Gal C, Sarkar S, Blaikley EJ, et al. A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. Nat Commun. 2014;5: 4091.

126. Zhu-Yr, Peery T, Peng-Tm, Ramanathan Y, Marshall N, Marshall T, et al. Transcription elongation factor p tefb is required for hiv 1 tat transactivation in vitro. Genes Dev. 1997;11: 2622–2632.

127. Sun S, Yang F, Tan G, Costanzo M, Oughtred R, Hirschman J, et al. An extended set of yeast-based functional assays accurately identifies human disease mutations. Genome Res. 2016;26: 670–680.

128. Youdell ML, Kizer KO, Kisseleva-Romanova E, Fuchs SM, Duro E, Strahl BD, et al. Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. Mol Cell Biol. 2008;28: 4915–4926.

129. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7: 539.

130. Schrödinger, LLC. The {PyMOL} Molecular Graphics System, Version~1.8. 2015 Nov.

131. Racusen LC, Monteil C, Sgrignoli A, Lucskay M, Marouillat S, Rhim JGS, et al. 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. 1997;129: 318–329.

132. Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, et al. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat Biotechnol. 2011;29: 697–698.

133. Meerbrey KL, Hu G, Kessler JD, Roarty K, Li MZ, Fang JE, et al. The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. Proc Natl Acad Sci U S A. 2011;108: 3665–70.

134. American Cancer Society. Cancer Facts & Figures 2016. Cancer Facts Fig 2016. 2016; 1–9.

135. Muglia V, Prando A. Renal cell carcinoma: histological classification and correlation with imaging findings. Radiol Bras. 2015;48: 166–174.

136. Seizinger BR, Rouleau GA, Ozelius LJ, Lane AH, Farmer GE, Lamiell JM, et al. Von Hippel-Lindau disease maps to the region of chromosome 3 associated with renal cell carcinoma. Nature. 1988;332: 268–9.

137. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. 2007;17: 877–885.

138. Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. Nature. 2003;423: 145–50.

139. Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, et al. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature. 2007;446: 572–6.

140. Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat Commun. 2016;7: 11485.

141. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

142. Gomez NC, Hepperla AJ, Dumitru R, Simon JM, Fang F, Davis IJ. Widespread Chromatin Accessibility at Repetitive Elements Links Stem Cells with Human Cancer. Cell Rep. 2016;17: 1607–1620.

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

144. Hacker KE. Investigating the role of SETD2 mutations and H3K36me3 loss in clear cell Renal Cell Carcinoma. Univ North Carolina Chapel Hill. 2014.

145. Lassmann T, Hayashizaki Y, Daub CO. TagDust - A program to eliminate artifacts from next generation sequencing data. Bioinformatics. 2009;25: 2839–2840.

146. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760.

147. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.

148. Chen K, Xi Y, Pan X, Li Z, Kaestner K, Tyler J, et al. DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res. 2013;23: 341–351.

149. Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol. 2011;12: R67.

150. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004;32: D493-6.

151. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31: 166–169.

152. Hacker KE, Fahey CC, Shinsky SA, Chiang Y-CJ, DiFiore J V., Jha DK, et al. 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. 2016;291: jbc.M116.739375.

153. Ryan MD, Drew J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. EMBO J. 1994;13: 928–33.

154. Donnelly MLL, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, et al. Analysis of the aphthovirus 2A/2B polyprotein “cleavage” mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal “skip.” J Gen Virol. 2001;82: 1013–1025.

155. de Felipe P. Skipping the co-expression problem: the new 2A “CHYSEL” technology. Genet Vaccines Ther. 2004;2: 13.

156. De Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD. E unum pluribus: Multiple proteins from a self-processing polyprotein. Trends Biotechnol. 2006;24: 68–75.

157. Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep. 2014;9: 1163–1170.

158. Schaft D, Roguev A, Kotovic KM, Shevchenko A, Sarov M, Shevchenko A, et al. The histone 3 lysine methyltransferase, SET2, is involved in transcriptional elongation. Nucleic Acids Res. 2003;31: 2475–2482.

159. Li J, Moazed D, Gygi SP. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. J Biol Chem. 2002;277: 49383–49388.

160. Shin H, Liu T, Manrai AK, Liu SX. CEAS: Cis-regulatory element annotation system. Bioinformatics. 2009;25: 2605–2606.

161. Kimura H. Histone modifications for human epigenome analysis. J Hum Genet. 2013;58: 439–445.

162. Wilkerson MD, Hayes DN. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26: 1572–1573.

163. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57: 289–300.

164. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12: 996–1006.

165. Zhu K, Lei P-J, Ju L-G, Wang X, Huang K, Yang B, et al. SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled alternative splicing. Nucleic Acids Res. 2016; 1–14.

166. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. 2006;20: 2922–36.

167. The_R_Development_Core_Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing,. 2016.

168. Gilbert TM, McDaniel SL, Byrum SD, Cades JA, Dancy BCR, Wade H, et al. 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. 2014;13: 2883–95.

169. Sadeh R, Launer-wachs R, Wandel H, Rahat A, Friedman N. Elucidating Combinatorial Chromatin States at Single-Nucleosome Resolution. Mol Cell. 2016;63: 1–9.

170. Liu KG, Gupta S, Goel S. Immunotherapy : incorporation in the evolving paradigm of renal cancer management and future prospects. Oncotarget. 2016;

171. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41: 955–961.

172. Li J, Duns G, Westers H, Sijmons R, Berg A van den, Kok K. SETD2: an epigenetic modifier with tumor suppressor functionality. Oncotarget. 2015;5.

173. Feng C, Sun Y, Ding G, Wu Z, Jiang H, Wang L, et al. PI3Kβ Inhibitor TGX221 Selectively Inhibits Renal Cell Carcinoma Cells with Both VHL and SETD2 mutations and Links Multiple Pathways. Sci Rep. 2015;5: 9465.

174. Weaver BAA, Cleveland DW. Decoding the links between mitosis, cancer, and chemotherapy: The mitotic checkpoint, adaptation, and cell death. Cancer Cell. 2005;8: 7–12.

175. Li Y, Zhang Z, Chen J, Huang D, Ding Y, Tan M, et al. VX680 / MK-0457 , a potent and selective Aurora kinase inhibitor , targets both tumor and endothelial cells in clear cell renal cell carcinoma. Am J Transl Res. 2010;2: 296–308.

176. Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, Shen H, et al. The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. Cancer Cell. 2014;26: 319–330.

177. Störkel S, Eble JN, Adlakha K, Amin M, Blute ML, Bostwick DG, et al. Classification of renal cell carcinoma: Workgroup No. 1. Union Internationale Contre le Cancer (UICC) and the American Joint Committee on Cancer (AJCC). Cancer. 1997;80: 987–989.

178. Speicher MR, Schoell B, du Manoir S, Schröck E, Ried T, Cremer T, et al. Specific loss of chromosomes 1, 2, 6, 10, 13, 17, and 21 in chromophobe renal cell carcinomas revealed by comparative genomic hybridization. Am J Pathol. 1994;145: 356–64.

179. Cheval L, Pierrat F, Rajerison R, Piquemal D, Doucet A. Of Mice and Men: Divergence of Gene Expression Patterns in Kidney. PLoS One. 2012;7: 1–12.

180. Nagy A, Wilhelm M, Sükösd F, Ljungberg B, Kovacs G. Somatic mitochondrial DNA mutations in human chromophobe renal cell carcinomas. Genes Chromosomes Cancer. 2002;35: 256–260.

181. Larman TC, DePalma SR, Hadjipanayis AG, Protopopov A, Zhang J, Gabriel SB, et al. Spectrum of somatic mitochondrial mutations in five cancers. Proc Natl Acad Sci U S A. 2012;109: 14087–14091.

182. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio S a JR, Behjati S, Biankin A V, et al. Signatures of mutational processes in human cancer. Nature. 2013;500: 415–21.

183. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012;149: 979–993.

184. Huang FW, Hodis E, Xu MJ, Kryukov G V, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013;339: 957–9.